

**Experimentelle Untersuchungen zur individuellen Variabilität
der menschlichen Sprechstimme**

Ingo Hertrich

Disbui. München
1987





Digitized by the Internet Archive
in 2026

Experimentelle Untersuchungen zur individuellen Variabilität
der menschlichen Sprechstimme

D i s s e r t a t i o n

der Fakultät für Biologie
der Ludwig - Maximilians - Universität München

vorgelegt von

Ingo Hertrich

aus München

Dezember 1966

1. Berichterstatter: Professor Dr. Dr. G. Ziegelmayer
2. Berichterstatter: Professor Dr. H. Cleave

Tag der mündlichen Prüfung: 22. Mai 1967

Vorbemerkungen

Die vorliegende Arbeit besteht im wesentlichen aus drei Teilen:

- (1) einem Literaturüberblick zum Thema menschliche Sprechstimme mit einer klassifizierten Bibliographie, die es ermöglicht, schnell einige Arbeiten zu bestimmten Teilgebieten (z. B. Sprecheridentifizierung) zu erhalten;
- (2) der Entwicklung einer zumindest teilweise (z. B. was die Untersuchung des Sprechrhythmus anbelangt) neuen Methode zur Extraktion sprecherbezogener Parameter aus dem Sprachsignal und der Durchführung mehrerer Voruntersuchungen zur Abschätzung technischer und methodischer Fehlerquellen,
- (3) einer Studie an einer größeren Stichprobe zur Erfassung korrelativer Zusammenhänge zwischen Sprechstimme, auditivem Perzept und allgemeinen Sprechereigenschaften (z. B. Körpergröße). In diesem Zusammenhang wird insbesondere der Geschlechtsdimorphismus untersucht sowie ein empirisch testbares Modell der nonverbalen Kommunikation konstruiert, das im wesentlichen auf dem Brunswick'schen Linsenmodell aufbaut.

Es sei hier erwähnt, daß diese Arbeit ohne die Unterstützung von vielen Seiten her nicht möglich gewesen wäre; mein Dank gilt allen Versuchspersonen, insbesondere den Beurteilern der Sprechproben, die sich in langen Sitzungen alle Tonbandaufnahmen anhören mußten, Herrn Professor Ziegelmayer für Hilfen bei der Themensuche, die Betreuung der Arbeit und die Bereitstellung eines Arbeitsraumes im Institut für Anthropologie und Humangenetik, Herrn Professor Cleve und allen Mitarbeitern des Instituts für ihre Anteilnahme und Ratschläge, Herrn Professor Tillmann (Institut für Phonetik und sprachliche Kommunikation) für Diskussionsbeiträge und die Ermöglichung von Voruntersuchungen an seinem Institut, Herrn Dr. Kotten für die Erstellung eines ersten Computerprogramms und damit wertvolle Hilfestellungen beim Einstieg in die digitale Signalverarbeitung, Herrn Dr. Schweisthal für methodische Anregungen, Einbeziehung meiner Fragestellung in Seminare des Instituts für Phonetik und sprachliche Kommunikation und insbesondere für die Überlassung von Sprechproben eines Stimmimitators zur statistischen Auswertung, Herrn Dr. Ruske (Lehrstuhl für Datenverarbeitung der TU München) für die Ermöglichung, die Sprechproben zu digitalisieren, die Bereitstellung von Magnetbändern und für Ratschläge bezüglich des technischen Vorgehens bei der digitalen Signalverarbeitung, Herrn Dr. Heß für die Einladung zu einem Seminarvortrag am Lehrstuhl für Datenverarbeitung und für Diskussionsbeiträge, allen Betreuern, Beratern und Mitarbeitern des Leibnitz-Rechenzentrums der Bayerischen Akademie der Wissenschaften, insbesondere Herrn Dr. Sarreither als Fachbetreuer und

Herrn Dr. Wiseman für Anregungen bezüglich statistischer Rechenverfahren, Herrn Dipl. Ing. Bernhard für die Übergabe digitaler Filterkoeffizienten und die Beratung im Umgang mit digitalen Filtern, außerdem allen, die mich durch ihr Interesse, ihre Kritik und ihre Anteilnahme motivierten, auf Ideen brachten und vor Fehlern schützten, insbesondere meinen Eltern, meiner Frau und meinen Kindern.

Inhalt

V2 . . . Vorbemerkungen

V4 . . . Inhalt

1 . . . 1. Einführung in die Thematik

1 . . . 1.1. Allgemeines

1 . . . 1.2. Die Natur des Sprachsignals

1 . . . 1.2.1. Physikalisches Modell der Spracherzeugung

2 . . . 1.2.2. Determination des Sprachsignals

3 . . . 1.2.3. Grenzen der Analysierbarkeit des Sprachsignals

4 . . . 1.3. Methoden der Analyse gesprochener Sprache

4 . . . 1.3.1. Auditive Methoden

4 . . . 1.3.2. Apparative Methoden

5 . . . 1.4. Literaturüberblick

6 . . . 1.4.1. Arbeiten mit anthropologischen Fragestellungen

6 . . . 1.4.1.1. Lebensalter und Stimme

7 . . . 1.4.1.2. Geschlechtsdimorphismus

8 . . . 1.4.1.3. Konstitution und Stimme

9 . . . 1.4.1.4. Vererbung und menschliche Stimme

11 . . . 1.4.2. Arbeiten aus verschiedenen angrenzenden Bereichen

11 . . . 1.4.2.1. Sprach- und Informationswissenschaften

13 . . . 1.4.2.2. Psychologie

18 . . . 1.4.2.3. Medizin

22 . . . 2. Fragestellung

26 . . . 3. Voruntersuchungen

26 . . . 3.1. Untersuchung zur Textunabhängigkeit des Langzeitspektrums

26 . . . 3.1.1. Material und Methode

28 . . . 3.1.2. Ergebnisse

30 . . . 3.2. Kasuistische Beobachtung an zwei eineiigen Zwillingspaaren

31 . . . 3.3. Erste Anhaltspunkte für einen Geschlechtsdimorphismus im Langzeit-
spektrum

35 . . . 3.4. Beobachtung bezüglich Sprachunabhängigkeit des Langzeitspektrums

35 . . .	3.5. Einfluß von Sprechtonhöhe und Lautstärke auf das Langzeitspektrum
35 . . .	3.5.1. Material und Methoden
38 . . .	3.5.2. Ergebnisse
47 . . .	3.6. Abhängigkeit des Langzeitspektrums von Mikrofon – und Raumakustik
49 . . .	3.7. Willentliche Beeinflußbarkeit des Langzeitspektrums bei Sprecher – Imitationsversuchen
55 . . .	3.8. Methode zur Erfassung des Langzeitspektrums im Frequenzbereich unterhalb 1000 Hz mit höherer Frequenzauflösung
55 . . .	3.8.1. Methode
57 . . .	3.8.2. Sprechlautstärke und Tonhöhe
60 . . .	3.8.3. Geschlechtsdimorphismus
62 . . .	3.9. Erfassung sprecherbezogener Parameter aus dem Sprechrhythmus
62 . . .	3.9.1. Intensitätsverlauf
64 . . .	3.9.2. Grundfrequenzverlauf
66 . . .	3.9.3. Berechnung von Rhythmusspektren aus Intensitäts – und Grundfrequenzverlauf
69 . . .	4. Studie an einer größeren Stichprobe von Sprechern zur Erfassung akustischer Eigenschaften und anthropologischen, psychologischen und auditiven Korrelaten
69 . . .	4.1. Datenmaterial
69 . . .	4.1.1. Akustische Analyse der Sprechproben
72 . . .	4.1.2. Anthropometrische Daten
73 . . .	4.1.3. Psychologische Daten
73 . . .	4.1.4. Auditive Beurteilungsdaten
73 . . .	4.1.4.1. Stimm – Selbstbeurteilung
73 . . .	4.1.4.2. Stimm – und Sprecher – Fremdbeurteilung
75 . . .	4.2. Geschlechtsdimorphismus
75 . . .	4.2.1. Varianzanalyse
77 . . .	4.2.2. Multivariate Ähnlichkeitsvergleiche
82 . . .	4.2.3. Automatische Geschlechtererkennung
83 . . .	4.2.3.1 Clusteranalyse
85 . . .	4.2.3.2. Diskriminanzanalyse
87 . . .	4.2.4. Auditive Geschlechtswahrnehmung
89 . . .	4.2.5. Sprechstimme und Andromorphie innerhalb eines Geschlechts
93 . . .	4.3. Auditive Stimmbeurteilung und akustische Korrelate

93	4.3.1. Selbst- und Fremdbeurteilung
100	4.3.2. Auditive Stimmbewertung und akustische Parameter
100	4.3.2.1. Faktorenstruktur der Stimmbewertungsattribute
102	4.3.2.2. Darstellung von Einzelkorrelationen zwischen
109	4.3.2.3. Abschätzung eines quantitativen Zusammenhangs zwischen auditiven und akustischen Parametern mit Hilfe multipler Regressionsfunktionen
113	4.4. Sprechstimme und sonstige Sprechereigenschaften
113	4.4.1. Empirisch testbares Modell der nonverbalen Kommunikation
120	4.4.2. Genauere Betrachtung der Korrelationen von Sprechereigenschaften mit einzelnen akustischen Parametern am Beispiel der Körpergröße
123	5. Diskussion
123	5.1. Methodik
123	5.1.1. Digitale Berechnung der Langzeitspektren
124	5.1.2. Parameterextraktion aus dem Sprechrhythmus
125	5.1.3. Datenreduktion durch Hauptkomponentenanalyse
126	5.1.4. Smallest Space Analysis
127	5.2. Diskussion der Ergebnisse im Hinblick auf verschiedene Fragestellungen
127	5.2.1. Sprecheridentifizierung
128	5.2.2. Sprechlautstärke und Tonhöhe
128	5.2.3. Geschlechtsdimorphismus
132	5.2.4. Auditive Stimmbewertung und akustische Korrelate
132	5.2.5. Nonverbale Kommunikation von Sprechereigenschaften
134	6. Zusammenfassung
137	Bibliographie
A1	Anhang
A1	Anthropometrische Merkmale
A3	Stimm – Selbstbewertungsfragebogen
A4	Stimm – Fremdbewertungsfragebogen
A5	Sprecher – Bewertungsfragebogen

1. Einführung in die Thematik

1.1. Allgemeines

Beim Menschen ist der akustische Kommunikationskanal durch eine hohe Komplexität ausgezeichnet. In erster Linie dient er zur Übermittlung von Sprache, einem ausgefeilten und zuverlässigen Informationsträger.

Im Gegensatz zur geschriebenen Sprache, die sich weitgehend auf den semantischen Gehalt des jeweiligen Textes beschränkt und die sich mit Hilfe einer Schreibmaschine sogar von verschiedenen Personen beliebig oft exakt reproduzieren läßt, stellt die gesprochene Sprache eine Abfolge singulärer Ereignisse dar, die zwar aufgrund ihrer Funktion als Textübermittler gewissen Restriktionen unterliegt, die aber dennoch von Sprecher zu Sprecher und von Situation zu Situation variiert. Diese Unschärfe kann einerseits als Rauschen angesehen werden, das die Informationsübermittlung stört, andererseits aber auch als ein zusätzliches Signal, das dem "eigentlichen" Sprachsignal überlagert ist.

Zur Analyse dieses nonverbalen Aspekts soll in dieser Arbeit ein Beitrag geleistet werden.

1.2. Die Natur des Sprachsignals

In diesem Kapitel werden in stark vereinfachter Form die wesentlichen Komponenten vorgestellt, die am Zustandekommen des Sprachsignals beteiligt sind.

1.2.1. Physikalisches Modell der Spracherzeugung

Ähnlich wie bei vielen Musikinstrumenten besteht der menschliche Sprechapparat im wesentlichen aus drei Teilen: einer Energiequelle, einem Vibrator bzw. Rauschgenerator und einem Resonator. Diese drei Teile können weitgehend unabhängig voneinander, gesteuert durch neuromuskuläre Prozesse, in der Zeit variiert werden.

(1) Als Energiequelle dient der Luftdruck, der in der Lunge erzeugt wird. Er wird hauptsächlich vom Zwerchfell und der Brustmuskulatur im Zusammenspiel mit Elastizitätseigenschaften des Brustraumes aufgebaut und variiert.

(2) Der Vibrator ist in der Regel bei stimmhaften Lauten durch die Stimmlippen im Kehlkopf realisiert. Sie öffnen und schließen sich und erzeugen so die Grundfrequenz der Sprechstimme. Diese Schwingung ist weder eine Impulsfolge noch sinusförmig; sie

enthält viele Obertöne. Angeregt wird sie durch den Luftstrom, reguliert durch die Hals- und Kehlkopfmuskulatur.

Stimmlose Laute entstehen dadurch, daß bei geöffneten Stimmlippen der Luftstrom eine Engstelle (z. B. Zunge/Gaumen oder obere Schneidezähne/Unterlippe) passiert, an der er sich "reibt" und eine Art Rauschen erzeugt. Durch Kombination von Kehlkopf- und Reibelaute entstehen stimmhafte Frikative (z. B. stimmhaftes "s"); bei Plosiven (z. B. "p") wird eine Artikulationsstelle (z. B. Lippen) nach Verschuß rasch geöffnet und gibt den "angestauten" Luftstrom frei (stimmhaft oder stimmlos).

(3) Den Resonator, dem der entscheidende Einfluß auf die Klangfärbung des jeweiligen Lautes zugesprochen werden kann, bilden im wesentlichen Rachen-, Mund- und Nasenraum mit seinen Nebenhöhlen. Rachen und Mundraum können durch die Zungen, Kiefer und Gesichtsmuskulatur stark beeinflusst werden, auf den Nasenraum wirkt in erster Linie der Schwellungszustand der Schleimhäute.

1.2.2. Determination des Sprachsignals

Wenn auch das physikalische Modell die mechanischen Prozesse aufzeigt, die bei der Spracherzeugung eine Rolle spielen, so geht aus ihm nicht hervor, durch welche Einflüsse der Sprechvorgang koordiniert und gesteuert wird. Aufgrund eines allgemeinen Vorverständnisses lassen sich die möglichen Einflußgrößen folgendermaßen aufgliedern:

(1) Relativ konstante Sprechereigenschaften:

- a) Morphologie des Sprechapparats
- b) feste neuronale "Verdrahtungen" und "Programme"
- c) Persönlichkeitseigenschaften
- d) Sprachgruppenzugehörigkeit des Sprechers (Muttersprache, Dialekt ...)
- e) Lebensalter (evtl. auch zu (3))
- f) Geschlecht

(2) Physikalische Umgebung:

- a) Luftfeuchtigkeit, Temperatur, Wind
- b) Raumakustik, Hintergrundgeräusch

(3) Zustand des Sprechers:

- a) biologisch/medizinisch (z.B. müde, erkältet, betrunken ...)
- b) psychologisch (z.B. Emotion, "Streß" ...)

(4) Absicht des Sprechers:

- a) verständliche Textübermittlung
- b) Wortwahl

- c) Setzen der Betonungen, Satzgliederung, "Unterton" (entsprechend der Situation und der Interessenlage des Sprechers, den Text auf eine bestimmte Art zu übermitteln)

Diese Aufgliederung, so schön sie im ersten Moment sein mag, enthält Begriffe aus sehr unterschiedlichen Disziplinen, die sich teilweise überlappen, die zum Teil nicht exakt definiert sind und die gewisse Schwierigkeiten erwarten lassen, wenn man das Sprachsignal analysieren will.

1.2.3. Grenzen der Analysierbarkeit des Sprachsignals

Das Ziel, alle Einflußgrößen beim Zustandekommen der gesprochenen Sprache auf ihren getrennten Beitrag zum Ganzen hin zu untersuchen wird durch folgende Hindernisse beeinträchtigt:

- (1) Die verschiedenen Komponenten sind größtenteils nicht unmittelbar zugänglich.
- (2) Sie unterliegen vermutlich sehr komplexen Wechselwirkungen untereinander.
- (3) Eine entscheidende Größe ist die Absicht des Sprechers. "Absicht" ist ein Begriff, der für uns zwar verständlich ist, weil auch wir hin und wieder Absichten haben, läßt sich aber weder quantifizieren noch operationalisieren. Wir können nur vermuten, wenn wir sehen, daß jemand, wenn er auf eine bestimmte Art spricht, in der Regel etwas bewirkt, was sich in unseren Augen für ihn als vorteilhaft herausstellt, daß er absichtlich so gesprochen hat. Das heißt, wir verwenden für die Erklärung eines Ereignisses (des Sprechakts) ein Ereignis, das erst später stattfindet (die Reaktion des Zuhörers). Unser Vorgehen ist daher teleologisch und kann zu einer kausalen Analyse im naturwissenschaftlichen Sinne nichts beitragen.
- (4) Auch Einflußgrößen, die zunächst nichts mit der Absicht des Sprechers zu tun haben, wie z. B. sein Alter oder Geschlecht, können nicht vollständig analysiert werden, weil es denkbar ist, daß z. B. ein Mann wie eine Frau sprechen will und es aufgrund der enormen willentlichen Einflußmöglichkeiten auf den Sprechvorgang nicht auszuschließen ist, daß er in seinem Vorhaben einen gewissen Erfolg hat.
- (5) Sprechen ist menschliches Verhalten. Die Frage nach einer vollständigen Analysierbarkeit des Sprechvorgangs berührt die Frage nach der Determiniertheit menschlichen Verhaltens überhaupt und kann hier wohl nicht geklärt werden.



1.3. Methoden der Analyse gesprochener Sprache

Das Sprachsignal selbst ist in seiner Rohform für die Analyse nicht gut zugänglich; man hätte nur eine unüberschaubare Datenfülle. Um eine Sekunde Sprache mit einem Auflösungsvermögen, das dem des menschlichen Ohrs entspricht, zu beschreiben, bräuchte man mindestens 30000 einzelne Zahlenwerte.

Es ist also erforderlich, entsprechend der Fragestellung, an der man interessiert ist, eine Reduktion der Daten vorzunehmen. Der menschliche Zuhörer kann dies, indem er seine Aufmerksamkeit nur auf bestimmte Aspekte lenkt und den Rest schnell wieder "vergißt". Für eine objektive Analyse braucht man ein Parameterextraktionsverfahren, das nur bestimmte Eigenschaften des Sprachsignals erfaßt.

Diese beiden Möglichkeiten, die auditive und die apparative, können auch kombiniert werden, indem entweder das Sprachsignal zunächst maschinell vorverarbeitet (z. B. gefiltert) und dann angehört wird oder indem es erst Zuhörern vorgespielt wird und dann aus deren Verhalten sekundäre Parameter abgeleitet werden.

1.3.1. Auditive Methoden

Neben dem einfachen Aufschreiben des gesprochenen Texts (Diktat) oder der Lautfolge (phonetische Transkription) und einer kasuistischen Beschreibung der Sprechereigenschaften gibt es detailliert ausgearbeitete Schemata, mit denen eine Sprechprobe ausgewertet werden kann. Man definiert sich Merkmale, Merkmalsausprägungen und/oder ein Klassifikationsraster und hält das Resultat auf einem dafür vorbereiteten Fragebogen fest.

Der Nachteil auditiver Verfahren ist ihre Subjektivität und ihre mangelnde Reproduzierbarkeit, der Vorteil ist ihre Einfachheit und die Tatsache, daß das menschliche Ohr im Zusammenhang mit dem menschlichen Gehirn ein noch weitgehend unübertroffenes Meßinstrument darstellt.

1.3.2. Apparative Verfahren

Eine objektive Sprachsignalanalyse kann mit analogen und mit digitalen Methoden betrieben werden. Auch Kombinationen sind denkbar, um einerseits die Exaktheit der digitalen Verfahren und andererseits die Schnelligkeit der analogen Verfahren zu nutzen. Nachdem technisch fast alle analogen Verfahren auch digital realisierbar sind, seien hier

analoge Verfahren nicht explizit beschrieben.

Die digitale Analyse besteht im wesentlichen aus drei Teilen: einem Wandler, der das Signal (die ankommende Schallwelle) in eine Zahlenfolge umwandelt, einem Rechenprogramm, das Parameter berechnet und einem Speicher – oder Ausgabemedium, in das die errechneten Daten geschrieben werden.

Beispiele für mögliche Parameter sind:

- Verlaufskurven:
 - Grundfrequenz
 - Intensität
 - Intensität einzelner Frequenzbänder
- Verlaufsstatistik:
 - Nulldurchgangsdichte
 - Varianz und Mittelwert von Intensität und Grundfrequenz
- Koeffizienten, die das Signalstück als Impulsantwort eines rekursiven Filters beschreiben (linear prediction analysis)
- Fourier – Spektrum und abgeleitete Größen:
 - Formanten
 - Langzeitspektrum
 - Cepstrum
- Autokorrelationsfunktion

Eine ausführlichere Darstellung findet sich bei Kunt & Hugli (1985).

1.4. Literaturüberblick

Nachdem die Untersuchung der Sprechstimme innerhalb der Anthropologie ein noch wenig bearbeitetes Thema darstellt, bietet sich an, zunächst einmal eine Bestandaufnahme zu machen, was an anthropologischen Arbeiten auf diesem Gebiet bereits existiert und dann weitgestreut verschiedene Wissenschaftsdisziplinen zu betrachten, aus denen zumindest methodische Anregungen zu erwarten sind.

Wenn auch diese Zusammenstellung keinen Anspruch auf Vollständigkeit erhebt, so soll sie doch einen Überblick verschaffen, der eine Grundlage für eigene experimentelle Untersuchungen darstellt.

1.4.1. Arbeiten mit anthropologischen Fragestellungen

1.4.1.1. Lebensalter und Stimme

Stimmentwicklung bei Kleinkindern

An älteren Arbeiten sind hier Einzelbeobachtungen an eigenen Kindern zu erwähnen, z. B. Fairbanks (1942), der bei seinem Sohn die Tonhöhenverteilung beim Hungerschreien (1/4 Stunde Schreienlassen!!) ermittelte oder Raffler Engel (1965), die die Laute ihres Sohnes auditiv beschreibt und hervorhebt, daß melodische Strukturen in der Entwicklung zeitlich vor artikulatorischen im Vordergrund stehen.

Eine Stichprobe von 20 Neugeborenen untersuchten Lieberman et al. (1971): Sie stellten eine Ähnlichkeit mit Affenlauten fest und schlossen daraus, daß aufgrund anatomischer Gegebenheiten Neugeborene nicht die Möglichkeit der Lautartikulation haben, wie sie beim Erwachsenen besteht (Rachenraum noch kaum vorhanden).

Longitudinalstudien für das erste Lebensjahr machten Maskarinek et al. (1981) und Kent und Murry (1982). Bei Maskarinek et al. stand die Frage im Vordergrund, wie sich die Hörfähigkeit der Kinder und Eltern auf die Lautproduktion auswirkt. Es zeigte sich, daß ein gewisses Spontanangebot unabhängig von Gehörtem in Erscheinung tritt, daß aber die Lautproduktionen bei tauben Kindern wieder zurückgehen, während sie bei hörenden allmählich sprachähnlicher werden. Die Störung war geringer, wenn lediglich die Eltern taub waren. Kent und Murry (1982) untersuchten deskriptiv die Entwicklung von Tonhöhe und Spektrum an einer Gruppe von 21 Kindern im Alter von 3, 5 und 9 Monaten.

Zum Geschlechtsdimorphismus von Kleinkindern gibt es zwei Arbeiten, eine von Bennett und Weinberg (1972), die andere von Ingrisano et al. (1980). Beide Studien weisen nach, daß das Geschlecht an der Stimme weit über dem Zufall erkannt wird; akustische Parameter hierfür konnten jedoch nicht isoliert werden. (Die Tonhöhe kann es nicht sein, da sie sich bei Buben und Mädchen kaum unterscheidet.)

Eine kulturvergleichende Studie bezüglich der "voice onset time" 9–12 Monate alter Kinder machte Enstrom (1982): Es zeigten sich weitaus größere Ähnlichkeiten in den Kindersilben englischer und arabischer Kinder als bei Erwachsenen in den entsprechenden Muttersprachen, was mehr für "nature" als für "nurture" bei den ersten Plapperlauten spricht. Außerdem wird noch erwähnt, daß im Schweizerdeutschen bezüglich der voice onset time genau das Kindermuster realisiert ist.

Man kann sagen, daß die Entwicklung der gesprochenen Sprache bei Kindern noch wenig untersucht ist; das vorhandene Wissen basiert auf Beobachtungen an wenigen Kindern und läßt noch kaum verallgemeinernde Schlußfolgerungen zu.

Die Stimme Jugendlicher

Die Tonhöhenentwicklung bei Jugendlichen untersuchten Hollien und Malcik (1962 und 1965) an Weißen und Neger in Amerika: Es wurden je sechs 10-, 14- und 18-Jährige jeder Gruppe untersucht und es zeichnete sich eine schwache Tendenz ab, daß Schwarze im Schnitt tiefere Stimmen haben und daß ihr Stimmbruch früher einsetzt.

Daß es nicht einen Stimmbruch gibt sondern mehrere, darunter auch "upward breaks", z. B. bei 10-Jährigen, beobachtete Fairbanks (1940).

Bezüglich spektraler und rhythmischer Muster in der Jugendlichenstimme scheint es noch keine größere Studie zu geben.

Stimmveränderungen mit dem Lebensalter bei Erwachsenen

Bezüglich der Sprechtonhöhe fanden Hollien und Shipp (1972) bei Männern von 20 bis 60 Jahren ein kontinuierliches Absinken mit zunehmendem Alter von 120 auf 110 Hz, bei 60- bis 90-Jährigen dann einen Anstieg auf über 140 Hz.

Das Höherwerden der Stimme im hohen Alter wird bestätigt durch Honjo (1980) und Mysak (1959).

Bei Frauen konnten Mc Glone und Hollien (1963) keine Grundfrequenzänderungen mit dem Alter nachweisen, während Honjo (1980) ein Absinken der Stimme im hohen Alter (über 70 Jahre) feststellte.

Weitere alterssensible Charakteristika finden sich im Zeitverlauf der Sprechstimme: Die Sprechgeschwindigkeit nimmt ab ca. 60 Jahren ab (Mysak 1959 und Ryan 1972), und bestimmte Artikulationsmuster (z. B. Vokallängen und voice onset time) weisen typische Veränderungen auf (Benjamin 1982 und Mysak 1959).

Daß sich das Alter in der Stimme überhaupt gut abbildet, zeigen auditive Tests, in denen das Lebensalter nach Vorspielen unbekannter Stimmproben sehr gut geschätzt wurde (Bonaventura 1935, Shipp und Hollien 1969 und Ryan und Burk 1974).

Inwieweit Stimmveränderungen auf anatomische Alterserscheinungen zurückzuführen sind, untersuchten Behrendt (1965) durch Autopsie an 84 Kehlköpfen und Honjo (1980) laryngoskopisch: Es zeigte sich unter anderem eine Altersatrophie der Stimmbänder.

1.4.1.2. Geschlechtsdimorphismus

Wie die Arbeiten von Bennett und Weinberg (1979) und Ingrisano et al. (1980) zeigen, besteht bereits bei Kleinkindern ein Geschlechtsdimorphismus in der Stimme (im Gegensatz zu früheren Behauptungen, s. Barth 1911)

Neben der Tatsache daß Frauen im Schnitt höhere Stimmen haben als Männer, wurden

auch andere geschlechtertrennende Aspekte gefunden. Coleman (1971) zeigte, daß bereits die Formanten (ohne F0) der "i" und "u" -Laute die Hörer in 80% der Fälle das Geschlecht erkennen lassen; Ähnliches gilt für manche stimmlosen Frikative (Ingemann, Set and Schwartz 1968) und geflüsterte Sprache (Lass et al. 1976 und Schwartz und Rine 1968).

Fall die männliche Stimme im Schnitt lauter ist als die weibliche, stellten Markel et al. (1972) fest. Sie konnten auch zeigen, daß in der dyadischen Kommunikation bei Gegengeschlechtlichkeit im Schnitt die Lautstärke größer ist als bei Gleichgeschlechtlichkeit der Gesprächspartner.

Mossen und Engbretson (1977) untersuchten unter Anderem den Geschlechtsdimorphismus der "glottal waveshape" mit Hilfe reflexionsloser Metallröhren (durch die angeblich der Einfluß der Resonanzräume eliminiert wird).

Zum Geschlechtsdimorphismus im Langzeitspektrum finden sich erste Ergebnisse bei Kiukaanniemi (1982) und Jacobshagen (1983).

Ein interessanter Bereich ist die Untersuchung der Maskulinität bzw. Femininität innerhalb eines Geschlechts. Terango (1966) stellte fest, daß bei Männern Stimmen als maskulin gelten, die eine niedrige Grundfrequenz aufweisen und diese wenig variieren.

Nach Brodnitz (1975) und Moses (1960) drückt sich in der Betonung bzw. Nivellierung des Geschlechtsdimorphismus der Zeitgeist bestimmter Epochen aus. Als Extremvariante bezüglich der menschlichen Stimme wäre hier die Kastratenstimme zu betrachten, in der wohl intersexuelle Archätypen realisiert werden und die zu bestimmten Zeiten sehr hoch im Kurs stand.

1.4.1.3. Konstitution und Stimme

Nachdem Steinwachs (1952) einen Zusammenhang zwischen Schreibdruck und Konstitution an ausgelesenen Kretschmer-Typen aufgezeigt hat, versuchten Grünewald und Zuberbier (1957) entsprechende Ergebnisse in der Sprechmotorik zu finden. Dabei gehen sie davon aus, daß die Psychomotorik etwas Ganzheitliches darstellt, das sich im gesamten Ausdrucksverhalten niederschlägt. Sie konnten bei Syndromen psychisch Kranker Parallelen zwischen Krankheitsbild, Schreibdruck-Verlauf und Sprechmotorik finden; auf eine größere Normalbevölkerungss Stichprobe wurden diese Studien jedoch nicht ausgedehnt, eventuell deshalb, weil damals die Meßmethoden für die Erfassung der normalen Variabilität noch zu grob waren.

Die auditiven Tests von Fay und Middleton (1940) zeigen, daß Leptosome und Pykniker weit über dem Zufall an der Stimme voneinander unterschieden werden können im

Gegensatz zu den Athletikern, für die es offensichtlich keine typischen Stimmcharakteristika gibt.

Moses (1941), der zum Zwecke der psychiatrischen Diagnose ein relativ kompliziertes auditives Auswertungsschema für die Untersuchung der menschlichen Sprechstimme entwickelt hat, kommt zu dem Ergebnis, daß sich die Konstitutionstypen nach Kretschmer und Sheldon in der Stimme widerspiegeln.

Weitere Hinweise für eine Korrelation zwischen Sprechstimme und Konstitution finden sich in den Zuordnungsexperimenten von Bonaventura (1935) und Allport und Cantril (1934) sowie bei Arndorfer (1965), der am Beispiel dreier erfolgreicher Redner auf konstitutionelle Korrelate in der Sprechstimme hinweist.

1.4.1.4. Vererbung und menschliche Stimme

Bernstein (1923) betrachtete die Tonlage der menschlichen Stimme als einfaches Mendel'sches Merkmal mit intermediärem Erbgang, wobei die reinerbigen Typen bei Männern Baß und Tenor sein sollten, bei Frauen Sopran und Alt. Baß und Sopran bzw. Alt und Tenor sollten jeweils Ausprägungen desselben Allels sein.

Mundinger (1951) und Hartlieb (1962) zeigten dann, daß dieses Modell nicht stimmt, daß die Tonlage in der Bevölkerung vielmehr normalverteilt ist, was, soweit es sich überhaupt um ein erbliches Merkmal handelt, für Polygenie spricht. Hartlieb (1962) stellte außerdem fest, daß sich für die Tonlagenbestimmung nur der tiefste erzeugbare Ton eignet, da das Ergebnis sonst zu stark von situationellen Einflüssen beeinträchtigt wird.

Daß zumindest bei Extremvarianten das Erbgut sich auf die Tonlage auswirkt, geht aus den Untersuchungen von Moran und Gilbert (1978) hervor. Die von ihnen untersuchten Down-Syndrom-Patienten hatten im Schnitt eine um 20 Hz erhöhte Grundfrequenz in ihrer Stimme. Kittel und Hecht (1977) untersuchten Kinder mit Katzenschrei-Syndrom und fanden Unterschiede zu Kindern mit normalem Chromosomensatz.

Eine bevölkerungsbiologische Beobachtung wurde von Darlington und Mather (1949), Thompson (1958), Brosnahan (1959) und Richter (1962) diskutiert. Es liegt die Beobachtung zugrunde, daß eine Korrelation besteht zwischen der Häufigkeit der Blutgruppe 0 und dem Gebrauch dentaler Frikative in Europa ($r = .46$). Es nimmt zwar niemand an, daß ein direkter ursächlicher Zusammenhang zwischen Blutgruppen und der Neigung, dentale Frikative zu gebrauchen, besteht; es wird auch nicht in Zweifel gezogen, daß sämtliche Bevölkerungen Europas in der Lage wären, sie zu produzieren; lediglich eine schwache Präferenz für den Gebrauch dieser Laute könnte genetisch programmiert

sein und sich in der jeweiligen Sprache niederschlagen. Die Blutgruppe 0 hätte dann den Stellenwert eines Markers. Thompson (1958) meint allerdings, daß im kulturellen Bereich bessere Erklärungen für die geographische Verteilung des Gebrauchs dentaler Reibelauts zu finden sind.

Einen Zusammenhang zwischen Stottern und Blutchemie versuchten Karlin et al. (1940) herzustellen, konnten aber keine signifikanten Unterschiede zwischen dem Blut von Stottern und dem von Nichtstottern finden.

Bezüglich rassistisch bedingter Unterschiede in der Stimme gibt es aus Amerika einige Beobachtungen: Moskowitz (1952) stellte fest, daß die Stimme schizophrener Neger "more gloomy" sei als die schizophrener Weiße.

Übereinstimmend berichten Hollien (1965) und Hudson (1982), daß Schwarze im Schnitt tiefere Stimmen aufweisen als Weiße.

Auch die Zuordnungsexperimente von Irwin (1977) und Lass (1978), in denen unbekannte Stimmproben von Weißen und Schwarzen auditiv weit über dem Zufall der "Rasse" richtig zugeordnet wurden, zeigen, daß sich die Zugehörigkeit zur schwarzen bzw. weißen Bevölkerungsgruppe in der Stimme ausdrückt.

Es gibt auch bereits einige Zwillingstudien an der menschlichen Stimme: Schilling (1936 und 1950) erstellte Einzelprotokolle von mehreren eineiigen Zwillingspaaren. Neben einer Reihe auditiver Parameter bezog er auch morphologische und psychologische Größen in seine Untersuchungen ein. Es konnte zwei Merkmalsgruppen voneinander trennen: In der ersten fand er hohe Übereinstimmungen bei eineiigen Zwillingen, und zwar in Kopfmaßen, Körpergröße, unterer Stimmgrenze und "primärem Kern der Klangfarbe"; die zweite Merkmalsgruppe, in der er häufig Unterschiede bei eineiigen Zwillingen fand, besteht aus "Visceralskelett" (= Nasenindex, Brustspielraum und Vitalkapazität), oberer Stimmgrenze, Stimmregister und "sekundären Klangbeimischungen".

Gedda et al. (1955 und 1960) zeigten, daß eineiige Zwillinge in ihrer Stimme so hohe Ähnlichkeiten aufweisen, daß sie oft die eigene Stimme nicht sicher von der Geschwisterstimme unterscheiden können, im Gegensatz zu zweieiigen die dies im allgemeinen sehr wohl können.

Keinen signifikanten Unterschied in den Ähnlichkeiten der Stimmen ein- und zweieiiger Zwillingspaare fand Ostwald (1962) bei Kindern; in Einzelfällen fand er auch bei Zweieiigkeit verblüffende Übereinstimmungen.

Das Klangspektrum von Stimmen eineiiger Zwillinge untersuchten Alpert und Kurtzberg (1963): Sie fanden große Ähnlichkeiten im Bereich zwischen 670 und 2000 Hz, während sich der Bereich unterhalb 400 Hz als variabel herausstellte.

1.4.2. Arbeiten aus verschiedenen angrenzenden Bereichen

1.4.2.1. Sprach- und Informationswissenschaften

Hier werden sehr unterschiedliche Bereiche zusammengefaßt; als Gemeinsamkeit könnte man eine gewisse Konvergenz auf das sicher sehr hoch gesteckte anwendungsbezogene Ziel einer automatischen Sprach- und Sprechererkennung bezeichnen.

Nachdem die menschliche Sprechstimme in ihrer Hauptfunktion zur Informationsübermittlung dient, ist die Analyse des sprachlichen Inhalts sehr eng mit der des sprecherspezifischen Signalaspekts verknüpft.

In der Phonetik ist man bemüht, eine text- oder aussagenbezogene Invarianz in der gesprochenen Sprache herauszuarbeiten. Dem menschlichen Zuhörer fällt dies relativ leicht, einerseits weil er den Text verstehen muß und es daher von klein auf gewohnt ist, auf ihn zu achten, andererseits weil er den Text versteht und dadurch aufgrund von Redundanzeigenschaften des Signals schlechtverständliche Teile ergänzen kann.

Der geübtere Zuhörer kann auch gesprochene Sprache transkribieren, ohne den Text zu verstehen, indem er, wohl hauptsächlich durch Introspektion, gelernt hat, auf die der Lautproduktion zugrundeliegenden Artikulationsbewegungen des Sprechers zurückzuschließen. Er stellt sich einfach vor, was er selbst machen müßte, um die gehörten Laute hervorzubringen. Das theoretisch denkbare akustische Kontinuum wird hierfür in Kategorien eingeteilt, über die erstaunlicherweise ein weitgehend sprachgruppenübergreifender Konsens besteht.

Untersuchungen wie die von Eimas (1985) an 4 Monate alten Kindern lassen vermuten, daß an diesem Prozeß angeborene Perzeptionsmuster beteiligt sind. Im Tierreich sind solche Muster bei verschiedensten Arten (z. B. Frösche, Vögel, Affen ...) bekannt.

Clauser (1971) nennt auch Hinweise dafür, daß durch pränatale akustische Wahrnehmungen bereits eine Vorstellung von sprachähnlichen Lauten erworben wird, die dann, sobald der Artikulationsapparat dazu in der Lage ist, sogar imitiert werden können.

Wesentlich problematischer als die auditive Analyse ist die objektive akustische Analyse der gesprochenen Sprache. Ein entscheidender Schritt hierfür war die Entwicklung des Sonagrammen, mit dem man die Schallwelle, die für sich genommen einen fast unüberschaubaren Komplex darstellt, in übersichtlicher Weise sichtbar macht (Potter, Kopp und Kopp 1947). Die zweidimensionale Signalstruktur (Schalldruck, Zeit) wird dreidimensional dargestellt (Frequenz, Intensität und Zeit), wobei impliziert wird, daß die Reduktion des Signals auf eine Abfolge annähernd stationärer Zustände (Die für eine

Frequenzanalyse erforderliche Integrationskonstante bewirkt eine zeitliche "Verschmierung") und die Eliminierung der Phasenbeziehungen keinen wesentlichen Informationsverlust darstellen.

Mit etwas Übung ist es auch tatsächlich oft möglich, den Text aus Sonagrammen zu "lesen", allerdings nicht, ohne wieder auf das gute Interpretationsvermögen des menschlichen Verstandes zurückzugreifen. Sprachlaute erkennt man an den Formanten (charakteristische Maxima im Frequenzspektrum stimmhafter Laute), an zeitlichen Mustern (Plosive und Lautübergänge) und an der Grobverteilung der spektralen Intensität (Frikative). (Siehe z. B. Halle et al. 1957, Fant 1973, Blumstein und Stevens 1979 und 1980.)

Auch Unterschiede von Mensch zu Mensch sind im Sonagramm sichtbar; daher wurden Sonagramme auch zur Sprecheridentifizierung verwendet (Kersta 1962). Allerdings zeigt sich, daß sie hierfür schlechter geeignet sind als akustische Darbietungen (Stevens 1966) und daß ihre forensische Anwendung als Beweismittel umstritten ist (Boit et al. 1973 und Black et al. 1973).

Die Entwicklung neuer Techniken ermöglichte dann eine gezieltere Darstellung von Parametern aus dem Sprachsignal; als Ziel steht die automatische Sprach- und Sprechererkennung im Vordergrund. Diese beiden Komplexe sind insofern eng miteinander verknüpft, als einerseits ein Sprecher ohne Sprache nicht als solcher identifiziert werden kann und andererseits die Sprache von jedem Sprecher unterschiedlich gesprochen wird.

Die Sprechererkennung machte zunächst große Fortschritte; als ein gut geeigneter Merkmalskomplex stellte sich das Langzeitspektrum heraus. Durch zeitliche Mittelung der spektralen Intensität über einen hinreichend langen Zeitraum (erforderlich sind mindestens ca. 15 s) wird der sprachliche Gehalt weitgehend eliminiert, während sich sprecherspezifische Eigenschaften herauskristallisieren. Durch Vergleich mit Hilfe eines Ähnlichkeits- oder Distanzmaßes (z. B. Korrelationskoeffizient, Euklidischer Abstand oder χ^2) und der Anwendung statistischer Methoden (z. B. Diskriminanzanalyse) kann dann die Sprechprobe dem Sprecher zugeordnet werden. (Smith 1962, Pruzansky 1963, Hargreaves und Starkweather 1963, Clarke und Becker 1969, Doherty 1976, Hollien und Majewski 1977, Bunge 1977)

Je nach Versuchsanordnung, Sprechproben- und Sprecherstichprobe wurden mit diesem Verfahren Sprechererkennungsraten von 80 – 100% erzielt.

Andere Langzeitstrukturen wie Langzeitautokorrelogramm (Goydke 1974), Korrelationsmatrizen von Spektralwerten (Li und Hughes 1974) oder "Sprecherkeulen" (Ungeheuer 1977) wurden ebenfalls zur Sprecheridentifizierung verwendet.

Versucht man, Sprecherunterschiede auch bezüglich sprachlicher Detailstrukturen zu analysieren, so rückt das Problem der Sprechererkennung und –charakterisierung wieder sehr nah an das der Spracherkennung heran. Man kann sich fragen: Worin liegt der Unterschied in der akustischen Realisierung, wenn zwei Sprecher "dasselbe" sagen? Es geht also um die interindividuelle Varianz sprachhomologer Segmente. Nachdem bis heute noch die vollautomatische Isolierung und Klassifikation solcher Segmente mit Schwierigkeiten verbunden ist, ist man auf eine auditiv/visuell – manuelle Vorverarbeitung angewiesen, wenn man sich nicht auf die Untersuchung sehr kurzer Äußerungen beschränkt.

Wie die Arbeiten von Luck (1969), Doddington (1971), Lummis und Rosenberg (1971), Hair und Rekieta (1972), Su, Li und Fu (1974), Joussein (1982) und Li und Wrench (1982) zeigen, ist es durchaus auch möglich, mit kurzen Sprachsegmenten oder –äußerungen gute Sprecher – Identifikationsraten zu erzielen.

Eine weitere Möglichkeit, Sprecherspezifitäten zu beschreiben, ist die Verwendung von Sprecher – Adaptionparametern aus Systemen der automatischen Spracherkennung (Jaschul 1982). Da diese Größen aber unter dem funktionalen Aspekt einer Verbesserung der Spracherkennung ermittelt werden, sind sie, für sich betrachtet, schwierig zu interpretieren.

1.4.2.2. Psychologie

In der Psychologie hat man sich schon seit längerem mit der Sprechstimme beschäftigt. Aufgrund der Tatsache, daß die Psyche selbst rein hypothetisch und nicht unmittelbar zugänglich ist, ist man gezwungen, aus dem, was an jemandem beobachtbar ist, auf sie zurückzuschließen. Die Stimme als menschliches Ausdrucksverhalten bietet sich hier natürlich an. Da ein Sprecher seine Aufmerksamkeit meist auf den verbalen Gehalt seiner Lautproduktionen lenkt, resultiert der nonverbale Aspekt der gesprochenen Sprache zu einem großen Teil aus dem Ablauf unbewußter Prozesse und stellt daher eventuell einen zuverlässigeren Indikator für das Innere des Sprechers dar als das, was er sagt. (Ein umfangreiches Literaturverzeichnis zu diesem Thema findet sich bei Scherer 1982.)

Studien zur Psychologie der menschlichen Sprechstimme lassen sich im wesentlichen in zwei Komplexe aufgliedern: Der erste besteht darin, die Stimme mit relativ konstant bleibenden Persönlichkeitseigenschaften (z. B. Extrovertiertheit) in Verbindung zu bringen, der zweite in der Erfassung eher temporärer Zustände des Sprechers (z. B. situative Anpassungen, Emotionen).

Sprechstimme und Persönlichkeit:

Will man die Stimme zur Persönlichkeit in Bezug setzen, so braucht man erst einmal ein Verfahren, um die Persönlichkeit des Sprechers, was auch immer man darunter verstehen mag, zu ermitteln. Neben kasuistischen Einzelbeurteilungen wurden hierfür vor allem Fragebögen verwendet. Auf die Problematik einer objektiven Persönlichkeitsdiagnostik soll hier allerdings nicht eingegangen werden; es sei einfach angenommen, daß die von verschiedenen Autoren beschriebenen Persönlichkeitsdimensionen existieren und sich "messen" lassen.

Bezüglich der Merkmalsextraktion aus der Sprechstimme taucht das Problem auf, daß es kaum einfach zugängliche Parameter gibt, die sich monokausal auf die Persönlichkeit des Sprechers zurückführen lassen. Gegenüber der Sprecheridentifizierungs-Fragestellung, bei der es genügt, *i r g e n d w e l c h e* Eigenschaften zu isolieren, an denen man einen Sprecher erkennen kann, ist hier nach ganz *s p e z i e l l e n* Eigenschaften gefragt, die typisch sind für ganz spezielle *a n d e r e* Eigenschaften des Sprechers.

In der Mehrzahl der auf diesem Gebiet durchgeführten Studien wurde mit auditiven Stimmkriterien gearbeitet, deren akustisches Korrelat noch wenig zugänglich ist.

Sapir (1927) unterschied fünf Ebenen der Analyse: Stimme als solche, Sprechdynamik, Aussprache, Vokabular und Stil.

Eine Liste von 21 Stimmkriterien erstellte Moses (1942). Es tauchen dabei Begriffe auf wie "Pathos", "Exactness", "Rhythmic prevalence" ... , über deren Bedeutung es zwar unter verschiedenen Zuhörern einen gewissen Konsens gibt, die aber kaum exakt definiert werden können.

Ein noch feiner ausdifferenziertes auditives Begriffsschema entwickelte Fährmann (1967, 1973 und 1977), das auch in der Praxis (Psychodiagnostik, Kriminalistik) angewendet wurde.

Fay und Middleton (1939, 1940, 1941, 1942 und 1943) führten umfangreiche Untersuchungen mit folgendem Ansatz durch: Sie bildeten nach verschiedenen Kriterien (z. B. Persönlichkeitstypen, Beruf ...) Gruppen von Sprechern und ließen dann Zuhörer nach Tonaufnahmen die Gruppenzugehörigkeit schätzen.

Das Problem der Parameterextraktion aus der Sprechstimme wurde somit umgangen; die Zuhörer mußten lediglich die Stimme der Sprecher ganzheitlich auf sich wirken lassen. Lagen die Schätzungen über dem Zufall, so konnte dies als ein Hinweis betrachtet werden, daß das jeweilige Merkmal in der Stimme oder Sprechweise zum Ausdruck kommt.

Gut geschätzt wurden die sechs Spranger-Typen ("theoretic", "economic", "aesthetic", "social", "political", "religious") und manche Berufe (z. B. Priester und Lehrer); weniger

gut wurden die Sprecher bezüglich der Merkmale "introversion", "leadership" und "sociability" erkannt.

Zum Teil zeigte sich, daß die Übereinstimmung unter den Zuhörern sehr hoch war, obwohl die Schätzung falsch war, was dafür spricht, daß es sich um "stereotyped judgments" handelt. Ähnliche Beobachtungen finden sich auch bei Kramer (1964).

Nach Scherer (1970 und 1978) könnte es allerdings durchaus möglich sein, daß diese Stereotype auf tatsächlichen statistischen Zusammenhängen beruhen, indem sich in den Zuhörern aufgrund individueller Vorerfahrungen und/oder kultureller Tradierungen "kognitive Inferenzkonstrukte" gebildet haben bezüglich Stimme und Persönlichkeit.

Die Beobachtung von Seip (1946), daß Zuhörer, die Stimmen beschreiben sollen, hierfür eher ein allgemein persönlichkeitscharakterisierendes Vokabular gebrauchen als ein auf die Stimmakustik oder die Artikulation bezogenes, spricht dafür, daß es bereits vorgeformte Perzeptionsmuster gibt, die die akustische Darbietung selbst gar nicht ins Bewußtsein dringen lassen sondern nur das, was für den Zuhörer "dahinter steht".

Solche komplexen Muster aber werden sicherlich nicht gebildet, wenn ihnen keine Funktion zukommt; sie könnten Teil eines nonverbalen Kommunikationskanals sein. Waldert-Lauth und Scherer (1983) wenden eine modifizierte Version des Brunswickschen Linsenmodells (Brunswick 1956) zur Beschreibung eines solchen Kommunikationskanals an: Eine psychische Eigenschaft wird mit Hilfe "distaler Hinweisreize" signalisiert; der Beobachter nimmt diese Reize als "proximale Perzepte" auf und attribuiert aufgrund kognitiver Inferenzregeln die entsprechende Eigenschaft.

Das Modell ermöglicht nun, die distalen Hinweisreize einerseits zu den tatsächlichen Eigenschaften, andererseits zu den proximalen Perzepten und der Attribution in Bezug zu setzen und so "Pfade" zu verfolgen, über die der nonverbale Kommunikationsprozeß läuft. Hierfür ist es erforderlich, die "black box" der gesamten funktionalen Einheit durch geeignete Operationalisierungen (z. B. Fragebögen, akustische Messungen, physiologische Messungen) transparent zu machen.

In einer Untersuchung von Scherer (1972) konnte bei einer deutschen Stichprobe die Eigenschaft "Dominanz" und bei Amerikanern die Eigenschaft "Extrovertiertheit" signifikant von Zuhörern textunabhängig (content masked) aus Sprechproben geschätzt werden. Erste Anhaltspunkte für Inferenzregeln, die der Attribution "extrovertiert" zugrundeliegen, finden sich bei Scherer (1978). (Siehe Abbildung 54, Seite 114)

Ohala (1970) beobachtete, daß die Eigenschaft "dominance" eine negative Korrelation zur mittleren Grundfrequenz der Sprechstimme aufweist.

Brown et al. (1973 und 1974) untersuchten den Einfluß von Grundfrequenz, Varianz der

Grundfrequenz und Sprechgeschwindigkeit auf die Perzeption der Eigenschaften "competence" und "benevolence" mit natürlichen und manipulierten (künstliche Veränderung der Grundfrequenz und der Sprechgeschwindigkeit) Stimmen und fanden folgende Effekte:

Beschleunigung der Sprechproben senkt die "benevolence", Verlangsamung senkt die "competence", Anheben der Grundfrequenz oder Vermindern ihrer Varianz senkt beides. Manipulierte Stimmen wurden allgemein negativer beurteilt als natürliche.

Phillis (1970) bringt Hinweise dafür, daß Untersuchungen an Kindern unter 10 Jahren geeignet sind, nonverbale Aspekte der Übermittlung von Persönlichkeitseigenschaften zu studieren, weil Kinder besonders sensibel seien für "nonverbal cues".

Sprechstimme und momentaner Ausdruck:

Temporäre psychische Einflußgrößen auf das sprachliche und stimmliche Ausdrucksverhalten eines Sprechers gliedern sich im wesentlichen in die drei Kategorien Emotion, Situation und Intention. Sie unterliegen komplexen Wechselwirkungen, was einerseits eine Analyse erschwert, wenn z. B. Emotionen absichtlich verdeckt werden, andererseits aber auch einen experimentellen Zugang erleichtert, indem z. B. Emotionen "gespielt" werden können.

Da es sich auch hier in der Regel um einen nonverbalen Kommunikationsprozeß handelt, kann sich eine Untersuchung entweder auf das Signal selbst richten oder auf den Perzeptionsprozeß beim Zuhörer. Letzteres hat den Vorteil, daß auch die Wirkung systematisch variiert synthetisierter Signale untersucht werden kann.

Bereits Rutz (1922) beschrieb für Schauspieler und Sänger einen Zusammenhang zwischen Gesamtsituation und Lautproduktion. Insbesondere die Körperhaltung wird hier betont, vermutlich deshalb, weil auch sie ein Element des Ausdrucksverhaltens darstellt, das aber willentlich leichter beeinflussbar ist als die Stimme selbst und daher zur "Einstimmung" des Lautproduzenten dienen kann.

Skinner (1935) fand bei der Untersuchung der Gefühle "happiness" und "sadness" Parallelen in Stimme, auditiver Selbst- und Fremdbeurteilung und Hautwiderstandsmeßwerten.

Fairbanks (1941) und Williams und Stevens (1972) konnten zeigen, daß von Schauspielern gespielte Emotionen von den Zuhörern über dem Zufall erkannt werden.

Es wurden auch Versuche unternommen, die akustischen Hinweisreize, die die Zuhörer

hierfür benutzen, einzugrenzen, indem das Signal entsprechend vorverarbeitet wurde (z. B. "random splicing" oder Tiefpaßfilterung). In erster Linie sollte dabei der Text unkenntlich gemacht werden (Scherer 1972).

Nach Starkweather (1956), Soskin (1961), Friedhoff et al. (1964) und Scherer et al. (1972) dient hauptsächlich der Frequenzbereich unterhalb ca. 600 Hz der Übermittlung von Gefühlszuständen. Insbesondere der Grundfrequenz – und Intensitätsverlauf könnte hier eine Rolle spielen (siehe z. B. Lieberman et al. 1962 und 1964, Huttar 1968, Constanzo 1969, und Scherer 1974a).

Ein Naturexperiment (ein Radiosprecher, der Zeuge einer unerwarteten Katastrophe wurde) zeigte allerdings, daß extreme Gefühlsveränderungen auch in höheren Spektralbereichen zum Ausdruck kommen (Williams und Stevens 1972).

Ein experimenteller Ansatz zur Untersuchung von echten (im Gegensatz zu gespielten) Emotionen besteht darin, auf Versuchspersonen einen Stimulus wirken zu lassen, der den jeweiligen Gefühlszustand hervorrufen soll.

Friedhoff et al. (1962) untersuchten so die Äußerung "no" als Antwort auf ein Emotionen weckendes Stimulus – Wort (z. B. "kiss"), das unter Nonsense – Wörtern eingestreut war. Ein 80 – dB – Störton sollte eine akustische Kontrolle der Antwort durch die Versuchspersonen verhindern. In Abhängigkeit vom Stimulus waren markante Energie – Umverteilungen im Klangspektrum bis 5000 Hz zu beobachten.

Ekman et al. (1976) spielten Versuchspersonen Filme vor, die positive oder negative Gefühle hervorrufen sollten. Sie gaben die Anweisung, bei nachfolgenden Interviews negative Gefühle zu verbergen. Beim Täuschen war ein Anstieg der Grundfrequenz zu beobachten, der aber von Zuhörern nur erkannt wurde, wenn das Signal durch Filterung textunverständlich gemacht wurde.

Pollack et al. (1960) konnten eine gute auditive Erkennung von gespielten Gefühlen auch unter erschwerten Bedingungen (hoher Geräuschpegel, geflüsterte Sprache) nachweisen. Weitere Literatur zu diesem Thema findet sich bei Scherer (1970b).

Zur Sprechsituation gibt es erst wenige Studien. Gregory und Hoyt (1982) untersuchten, inwieweit sich Gesprächspartner im Sprechrhythmus aneinander angleichen, indem sie die Umhüllende der Signals mit einer Samplingfrequenz von 80 Hz digitalisierten und Fourier – transformierten, konnten aber nur sehr schwache Effekte nachweisen.

Universale Stimmeigenschaften, durch die "politeness" zum Ausdruck kommt, beschreiben Brown und Levinson (1978).

Auditive Tests zur soziokulturellen und situationellen Variabilität des gesprochenen Französisch bezüglich Rhythmus, Pausensetzung, Intonation und Betonungssatzung führte

Lucci (1978) durch.

1.4.2.3. Medizin

Medizinische Untersuchungen an der menschlichen Stimme wurden meist mit konkreten Zielsetzungen bezüglich Pathologie der Stimme, Diagnostik mit Hilfe der Stimme oder Tauglichkeit der Stimme (z. B. bei Berufssängern) durchgeführt. Die angewendeten Untersuchungsmethoden sind auf die jeweiligen Fragestellungen zugeschnitten und lassen sich nur bedingt auf andere Fragestellungen übertragen. Da die Medizin sich meist nur mit (pathologischen) Extremvarianten beschäftigt, sind erstens die Methoden oft zu grob für eine Beschreibung der "normalen" Variabilität und zweitens die untersuchten Stichproben (z. B. Patienten in einem Krankenhaus) nicht repräsentativ für die Bevölkerung.

Die im Folgenden zitierten Arbeiten sind nur wenige Einzelbeispiele aus dem großen Feld medizinischer Untersuchungen an der menschlichen Stimme.

Pathologie der Stimme:

Die meisten Stimmabnormalitäten sind durch Störungen der im Kehlkopf produzierten Grundschwingung bedingt. Es finden sich hier Begriffe wie "breathy voice quality" (Ladefoged 1984), "vocal fry" (Coleman 1963, Wendahl et al. 1963, Whitehead und Emanuel 1974 und Klingholz und Martin 1983b), "harsh voice" (Wendahl 1963) und "hoarseness" (Yumoto 1984).

Akustisch drücken sich diese Eigenschaften dadurch aus, daß die Schwingung nicht gleichmäßig periodisch abläuft sondern starke Amplituden- (= "shimmer") oder Frequenzschwankungen (= "jitter") von Periode zu Periode aufweist.

Moser (1984) versuchte, durch mathematischen Vergleich benachbarter Schwingungen ein objektives Maß für die Stimmreinheit zu erhalten.

Yumoto et al. (1984) berechneten aus ca. 600 Einzelschwingungen ein Maß, das die Eigenschaft "hoarseness" beschreiben soll.

Ein multidimensionales Klassifikationsprogramm für Stimmabnormalitäten erstellten Murry et al. (1974).

Ein weiteres Thema ist die Untersuchung von Ersatzvibratoren bei mangelnder oder fehlender Kehlkopffunktion (z. B. nach Kehlkopfoperationen). Klingholz und Martin (1982) nennen hier die Taschenbandstimme, die Ösophagusstimme, die Pharynxstimme und die Tracheo – Ösophagealfistel.

Die Veränderung der Stimme nach Eingliederung von Gaumenplatten in Bezug auf das Frequenzspektrum untersuchten Altmann et al. (1981).

Inwieweit eine Mandeloperation die Stimme verändert, untersuchten Arndt (1963) und Wallner et al. (1968); beide kommen zu dem Ergebnis, daß Veränderungen nur kurz nach der Operation zu beobachten sind, vermutlich durch eine vorübergehende schmerzbedingte Schonhaltung.

Den Einfluß verschiedener Drogen (Droperidol, Meperidin, Skopolamin und Kokain) auf die Kehlkopffunktion untersuchten Fishman et al. (1971). Sie konnten bezüglich Grundfrequenz und "vocal fry" keine signifikanten Effekte nachweisen.

Hargreaves (1961) und Starkweather und Hargreaves (1964) konnten auch keine konsistenten Effekte von Natriumpentobarbital auf die Grundfrequenz und das Klangspektrum nachweisen; im Zeitverlauf beim Sprechen ("hesitation and pausing") zeigten sich jedoch Veränderungen.

Die Wirkung androgener Hormone auf die weibliche Stimme beschreibt Damste (1964 und 1967). Die Veränderung war ein irreversibles Absinken der Grundfrequenz. Berendes (1968) betont, daß die hierfür erforderliche Dosis individuell sehr unterschiedlich ist. Ähnliche Stimmveränderungen wurden auch bei Schwangerschaft (Imre 1951 und Deuster 1977) und nach der Menopause (Gilbert und Weismer 1974) beobachtet.

Imre (1968) zeigt, daß einige Hormondrüsen bei Dysfunktion Stimmstörungen verursachen können.

Gilbert und Weismer (1974) fanden einen Unterschied in der Grundfrequenz von Raucherinnen und Nichtraucherinnen: Erstere hatten im Schnitt eine um ca. 10 Hz tiefere Stimme.

Diagnostik mit Hilfe der Stimme:

Nachdem die Stimme, insbesondere die Sprechstimme, ein wichtiges Ausdrucksmittel für die Psyche des Sprechers darstellt, wurden vor allem auf dem Gebiet der Psychiatrie Versuche unternommen, Zusammenhänge zwischen Krankheitsbildern und Sprechstimme aufzudecken.

Moskowitz (1952) charakterisierte auditiv die Stimme Schizophrener als "monotonous", "weak", "unsustained", "gloomy", "colorless" und "flat".

Bezüglich Sprechgeschwindigkeit und Varianz der Grundfrequenz fanden Saxman und Burk (1968) Unterschiede zwischen schizophrenen und "normalen" Frauen: Schizophrene sprachen langsamer (gelesener Text) und variierten die Grundfrequenz mehr.

Ostwald (1965) beschreibt eine Beobachtung an einem schizophrenen Bub: Je nachdem, ob der Bub Nonsens, stereotype Sätze oder normal redete, nahm das über ca. 4 Sekunden gemittelte Frequenzspektrum eine für den jeweiligen Zustand charakteristische Form an.

Eine weitere auditive Studie zur Diagnose der Schizophrenie nach der Stimme ist von Chevrie – Muller et al. (1978) beschrieben.

Clemmer (1980) benutzte verschiedene Maße aus dem Zeitverlauf der Sprechstimme (z. B. Silbenlängen und Sprechgeschwindigkeit) und bekommt mit Hilfe der Diskriminanzanalyse eine gute Trennung zwischen Schizophrenen und "Normalen".

Auch bei Depressiven wurden Auffälligkeiten an der Sprechstimme beobachtet: Zuberbier (1957) beschreibt parallele Veränderungen in der Sprechmotorik und im Schreibdruckverlauf (Verlangsamung, Monotonie und verringerte Intensität). Hargreaves und Starkweather (1964) und Hargreaves et al. (1965) fanden Auswirkungen der Depression auf das Klangspektrum, die durch auditive Beurteilungen bestätigt wurden; die Art der Veränderung war jedoch individuell unterschiedlich.

Die Epilepsie drückt sich nach Spitznagel (1959) im Schreibdruckverlauf und in der Sprechmotorik unter anderem durch Rhythmusstörungen, Verlangsamung und erhöhte Grundspannung aus.

Die Stimmen von Parkinson – Patienten untersuchte Canter (1963 und 1965) und fand eine erhöhte Grundfrequenz, eine erniedrigte Varianz der Grundfrequenz und eine verminderte Fähigkeit, willentlich die Stimme zu beeinflussen.

Bereits im Säuglingsalter scheint die Stimme als diagnostisches Hilfsmittel von Bedeutung zu sein: Raes et al. (1980) erstellten Spektrogramme von Baby – Schmerzscreien und fanden charakteristische Auffälligkeiten z. B. bei Hyperbilirubinämie, Störungen des Zentralen Nervensystems und Unterernährung.

Tauglichkeitsuntersuchungen an der Stimme:

Für Berufe, bei denen der Einsatz der Stimme eine bedeutende Rolle spielt (z. B. Schauspieler und Sänger) wird eine ärztliche Stimmuntersuchung empfohlen (Frank 1980). Deinse (1974) meint, daß auch Singschüler bezüglich ihrer optimalen Tonlage genau untersucht werden müssen, weil Fehlklassifikationen zu Schäden führen können.

Bei der Stimmfeldmessung wird der Bereich erzeugbarer Töne (Intensität und Grundfrequenz) ermittelt und in ein zweidimensionales Diagramm (Frequenz gegen Intensität) eingetragen (Gross 1980 und Klingholz und Martin 1983a). Zusätzlich kann der Bereich markiert werden, in dem die Stimme heiser klingt (Heinemann und Gabriel 1982). Sovijärvi (1965) bestimmte die Stimmkategorien mit Hilfe von Resonanzröhren von 26 – 28 cm Länge.

Hinweise bezüglich einer Abhängigkeit des Stimmumfangs von der Tageszeit beschreibt Frank (1980).

Den Einfluß von Menstruation und Schwangerschaft auf die Singstimme von Berufssängerinnen beschreiben Flach et al. (1968).

Die Durchschlagskraft der Sprechstimme kann man messen, indem man ermittelt, wie laut ein Störschall, den ein Zuhörer über Kopfhörer bekommt, sein darf, damit er einen Sprecher in genau definiertem Abstand gerade noch versteht (Essen 1951).

Die Belastbarkeit der Stimme testeten Seidner et al. (1981): Sie ließen die Sprecher 1/2 Stunde mit gleicher Lautstärke (bei Unterschreiten der Soll-Lautstärke leuchtet eine Kontrolllampe auf) lesen und führten anschließend Tests durch, um Ermüdungserscheinungen zu messen.

2. Fragestellung

Aufgrund der hohen Komplexität der Vorgänge bei der Sprachproduktion ist das Ziel einer vollständigen Analyse der beteiligten Komponenten noch in weiter Ferne. Für die vorliegende Arbeit wurden daher nur einige wenige Aspekte herausgegriffen, denen sicher dadurch eine gewisse Willkür oder Zufälligkeit anhaftet, als in ihre Auswahl die technische Durchführbarkeit der Untersuchungen, wie sie sich für den Autor zu Beginn der Arbeit darstellte, nicht unwesentlich einging.

Insbesondere von einer Analyse kausaler Zusammenhänge, auf die in der Regel in den Naturwissenschaften die größte Aufmerksamkeit gerichtet ist, wurde hier weitgehend aus folgenden Gründen abgesehen:

- (1) Der Sprachschall wird von einem Apparat erzeugt, der größtenteils aus Weichteilen besteht. Selbst mit Röntgenuntersuchungen läßt sich seine Morphologie als Funktion der Zeit beim Sprechen nicht mit hinreichender Genauigkeit beschreiben.
- (2) Die Beschaffenheit der Weichteile in Bezug auf Elastizität und Oberfläche läßt sich am lebenden Körper nicht ohne komplizierte Eingriffe ermitteln; totes Material ist aber schon zu stark verändert; außerdem sprechen Tote nicht mehr.
- (3) An der Sprachproduktion ist eine Vielzahl von Muskeln beteiligt, deren gesonderte Aktivität nur mit Hilfe intramuskulärer elektrischer Ableitungen aufgezeichnet werden kann, was Versuchspersonen nicht zuzumuten ist.
- (4) Den Muskelbewegungen sind neuronale Prozesse vorgeschaltet, die zum großen Teil vom Sprecher willentlich beeinflussbar sind und somit nicht als einfaches Resultat einer Ursache – Wirkungs – Kette verstanden werden können.

Sieht man von einer Analyse kausaler Zusammenhänge ab, so bietet sich zunächst eine deskriptive Betrachtungsweise an.

Da es um sprecherbezogene Eigenschaften des Sprachsignals geht, ist es hierfür erforderlich, aus dem Signal Parameter zu extrahieren, die, möglichst unabhängig vom semantischen Gehalt des Gesagten, Aussagen über den jeweiligen Sprecher zulassen.

Die sprecherbezogenen Eigenschaften lassen sich nun wieder unterteilen in situationell bedingte, momentane Eigenschaften und in solche, die bei einem Sprecher über längere Zeiträume hin weitgehend erhalten bleiben. Letzteren soll hier, als eine weitere Einschränkung der Fragestellung, in erster Linie die Aufmerksamkeit geschenkt werden.

Weiterhin sei an die zu extrahierenden Parameter die Anforderung gestellt, daß es sich nicht nur um irgendwelche Zahlenwerte handelt, sondern daß sie noch eine gewisse Anschaulichkeit, das heißt wenigstens noch andeutungsweise auditive Korrelate besitzen.

Da eine Verwandtschaft zur Sprecheridentifizierungs – Fragestellung besteht, liegt es nahe, methodisch dort Anregungen zu holen. In erster Linie sei hier das Langzeitspektrum, also das über einen längeren Zeitraum (z. B. 20 Sekunden) gemittelte Klangspektrum erwähnt, das gute Identifizierungsraten ermöglicht und das Eigenschaften beschreibt, die vom menschlichen Ohr auch wahrgenommen werden können. (Z. B. wäre es möglich, ein solches Klangspektrum mit Hilfe einer Filterbank oder der Fourier – Synthese Zuhörern akustisch darzubieten.)

Da aber im Langzeitspektrum jegliche Information über den zeitlichen Verlauf des Sprachsignals fehlt, können damit Sprechereigenschaften, die in der Rhythmik oder Prosodie, also in Strukturen des Zeitverlaufs liegen, nicht erfaßt werden.

Es sei daher auch noch nach Parametern gesucht, die über den Zeitverlauf Auskunft geben, wobei insbesondere das Problem der Textunabhängigkeit wieder in den Vordergrund rückt, da der Zeitverlauf sehr stark von der Laut – und Syntaxstruktur des Gesprochenen abhängt.

Zur Lösung dieses Problems bieten sich zwei Möglichkeiten an (zwischen denen in dieser Studie eine Kompromißlösung verwirklicht wurde): Eine besteht darin, von allen Sprechern denselben Text lesen zu lassen, die andere darin, sehr lange Signalstücke zu betrachten und rhythmusbezogene Parameter zu extrahieren, in der Hoffnung, daß sich Einflüsse des konkreten Texts herausmitteln. Die zweite dieser Möglichkeiten könnte (ähnlich wie bei Gregory 1982) dadurch verwirklicht werden, daß ein Langzeitspektrum über den groben Zeitverlauf erstellt wird, in das Signalstücke von mehreren Sekunden Dauer eingehen, die auf Periodizitäten im Bereich der Rhythmuswahrnehmung hin (ca. 0.3 – 10 Hz) untersucht werden. Als zugrundezulegendes Zeitsignal bietet sich der Intensitäts – oder der Grundfrequenzverlauf der Sprechprobe an.

Zur Ermittlung des Grundfrequenzverlaufs muß noch erwähnt werden, daß es hierfür noch kein allgemein etabliertes, vollautomatisch und fehlerfrei arbeitendes Verfahren gibt; ein solches muß, den Anforderungen der Fragestellung entsprechend, im Hinblick auf die gewünschte Frequenz – und Zeitauflösung neu konzipiert werden. Insbesondere der Umgang mit Sprechpausen und stimmlosen Frikativen stellt ein Problem dar.

Untersucht man jeweils dieselbe Textstelle bei verschiedenen Sprechern, so läßt sich auch die Sprechgeschwindigkeit als einfacher Parameter leicht ermitteln. Aus dem Rhythmusspektrum kann dann der Einfluß der Sprechgeschwindigkeit (soweit er linear ist) dadurch eliminiert werden, daß die Signalstücke aller Sprecher auf dieselbe zeitliche Länge gedehnt bzw. komprimiert werden.

Diese Langzeitstrukturen in Klang und Rhythmus geben sicher nur einen kleinen Teil der

sprecherspezifischen Eigenschaften im Sprachsignal wieder. Akustische Eigenschaften (z.B. Formantstrukturen) einzelner, nicht häufig vorkommender Sprachlaute sowie lautbezogene Artikulationsgewohnheiten (z. B. Aspiration bei Plosiven) werden nicht erfaßt. Dieser Mangel ist einerseits technisch bedingt, da die objektive Isolierung und Identifikation von Sprachsegmenten noch nicht fehlerfrei funktioniert, andererseits dadurch, daß entweder eine sehr große Datenmenge entstehen würde, über deren Relevanz dann nur mit noch viel größeren Sprecherstichproben Aussagen möglich wären, oder daß eine relativ willkürliche Merkmalsauswahl getroffen werden müßte, die die Ergebnisse vom Zufall schlecht abgrenzen läßt. Die Langzeitstrukturen sind nicht ganz so willkürlich; erstens haben sie sich (zumindest das Klangspektrum) bei der Sprecheridentifizierung als einigermaßen reproduzierbar herausgestellt; zweitens sind sie ein Kondensat über das gesamte Signal und daher vom Sprecher willentlich sicher nicht so leicht manipulierbar wie einzelne Sprachlaute.

Zusätzlich zur deskriptiven Erfassung von sprecherbezogener Eigenschaften in der gesprochenen Sprache soll nun versucht werden, korrelative Zusammenhänge mit sonstigen Sprechermerkmalen zu finden, z. B. mit Geschlecht, Lebensalter, Konstitution und Persönlichkeitsmerkmalen.

Solche Zusammenhänge sagen zwar noch nichts über kausale Beziehungen aus, sie können aber nutzbar gemacht werden, indem die Stimme als Prädiktor für andere Merkmale des Sprechers verwendet werden kann. Die Güte dieses Prädiktors läßt sich empirisch leicht überprüfen, so daß man nicht darauf angewiesen ist, auf leicht zu verfälschende statistische Signifikanztests zu vertrauen. Beispielsweise läßt sich ermitteln, zu welchem Prozentsatz das Geschlecht unbekannter Sprecher objektiv aus der Sprechstimme erkannt werden kann.

Zum Begriff der Konstitution sei noch erwähnt, daß dieser Begriff oft mit subjektiven Typognose-Methoden in Verbindung gebracht wird und zum Teil auf heftige Kritik gestoßen ist, insbesondere was die Interpretation psychophysischer Korrelate anbelangt. Im Rahmen dieser Arbeit sei unter Konstitution lediglich eine Menge von meßbaren anthropometrischen Merkmalen und aus ihnen mit Hilfe reproduzierbarer und eindeutig festgelegter Rechenverfahren (z. B. Hauptkomponentenanalyse) abgeleiteten Größen verstanden.

Für den Fall, daß sich tatsächlich Zusammenhänge zwischen Sprechstimme und sonstigen Sprechermerkmalen ergeben, kann man sich fragen, ob auch der menschliche Zuhörer in der Lage ist, diese Zusammenhänge wahrzunehmen. Diese Frage führt zur Aufstellung eines Modells der nonverbalen Kommunikation. Ungeachtet des Entstehungs-

mechanismusses könnten bestimmte Eigenschaften des Sprachsignals die Funktion haben, dem Zuhörer etwas über den Sprecher mitzuteilen. Ähnlich wie bei Waldert – Lauth und Scherer (1983) könnte man hier eine Version des Brunswick'schen Linsenmodells konstruieren, um diesen Aspekt zu beleuchten.

Zu diesem Zweck wurden Fragebögen entworfen (siehe Anhang), die von Zuhörern nach Anhören der Sprechproben auszufüllen sind und die Auskunft darüber geben sollen, welchen Eindruck vom Sprecher, seiner Stimme und seiner Sprechweise das akustische Signal im Zuhörer erweckt. Die Fragebogendaten können nun sowohl mit tatsächlichen Sprechereigenschaften verglichen werden als auch mit den Parametern, die die akustische Analyse liefert. Man bekommt so eine Antwort auf die Fragen:

- (1) Kann der Zuhörer zuverlässige Aussagen über einen Sprecher machen?
- (2) Welche Eigenschaften des Sprachsignals dienen hierfür als Indikatoren?
- (3) Geht der Eindruck, den der Zuhörer sich macht, parallel mit dem, den der Computer sich aufgrund multipler Regressionsbeziehungen macht?
- (4) Welche akustischen Eigenschaften hat ein Sprachsignal, das im Zuhörer einen bestimmten Eindruck vom Sprecher und seiner Stimme hinterläßt?

Zusammenfassend besteht die Fragestellung also aus folgenden Punkten:

- (1) Extraktion von sprecherbezogenen Parametern aus dem Sprachsignal,
- (2) Auffinden korrelativer Zusammenhänge zwischen Sprechstimme und sonstigen Sprechereigenschaften,
- (3) Wirkung des Sprachsignals auf den Zuhörer in Bezug auf Sprechereigenschaften,
- (4) Vergleich der Zuhörerattributionen mit tatsächlichen Eigenschaften und mit objektiv extrahierten akustischen Parametern.

Dem Punkt (1) kommt dabei das größte Gewicht zu, da die Extraktion akustischer Parameter aus dem Sprachsignal in diesem Umfang ein für die Anthropologie neues Gebiet darstellt und umfangreiche Voruntersuchungen erfordert. In erster Linie soll ein zuverlässiges methodisches Instrumentarium für künftige Studien erarbeitet werden.

Punkt (2) bis (4) seien mehr exemplarisch verstanden; die Auswahl der für die Korrelationsanalyse verwendeten Merkmale richtet sich in erster Linie nach der möglichst leichten Zugänglichkeit von Sprecher – Stichprobe (Biologiestudenten und – studentinnen) und Meßmethoden (Anthropometrie, Fragebögen).

3. Voruntersuchungen

Bevor eine größere Stichprobe von Sprechern in Bezug auf sprecherspezifische Merkmale untersucht werden kann, ist es erforderlich, um den Rechenaufwand in Grenzen zu halten, an kleineren Stichproben Voruntersuchungen durchzuführen, und so ein zuverlässiges Parameterextraktionsverfahren zu entwickeln. Insbesondere folgende Aspekte sollen dabei beleuchtet werden:

- Textunabhängigkeit und Sprachunabhängigkeit
- Intraindividuelle Stabilität der Merkmale
- Variabilität aufgrund von Raumakustik und Mikrofonanordnung
- Fehlerfreies Arbeiten des Rechenprogramms

3.1. Untersuchung zur Textunabhängigkeit des Langzeitspektrums

3.1.1. Material und Methode

Von 5 männlichen und 5 weiblichen Versuchspersonen im Alter von 20 – 30 Jahren wurde ein gelesener Text (die Äsop – Fabel "der Nordwind und die Sonne") auf Tonband aufgenommen (Tonbandgerät Uher universal 5000, Mikrophon Uher M155, Mund – Mikrophon – Abstand ca. 20 cm). Die akustische Auswertung wurde am Institut für Phonetik und sprachliche Kommunikation der Universität München durchgeführt; das Computerprogramm hierfür entwickelte Dr. K. Kotten, der mir dadurch den Einstieg in die digitale Signalverarbeitung ermöglichte.

Die Sprechproben wurden pegelangepaßt, wodurch zwar die Information über die Absolutlautstärke verloren geht, dafür aber der Dynamikbereich des Analog – digital – Wandlers besser genutzt werden kann.

Daraufhin wurden sie bei 10 kHz tiefpaßgefiltert und mit einer Samplingrate von 20 kHz und einem Wertebereich von 12 bit digitalisiert und auf Magnetband abgespeichert.

Jedem Text wurden zwei Proben von ca. 20 Sekunden Dauer entnommen, so daß insgesamt 24 Proben zur Verfügung standen.

Die Berechnung der Langzeitspektren wurde folgendermaßen durchgeführt: Das Zeitsignal wird in Stücken von ca. 25 ms (=512 Samples) schrittweise verarbeitet. Von jedem Einzelstück wird nach Passieren einer Pegelschwelle (Sprechpauseneliminierung), Differenzierung und Bewertung mit einem Hamming – Fenster ein Amplitudenspektrum von 256 Werten (40 – 10000 Hz) mit Hilfe der schnellen Fourier – Transformation berechnet. Die so gewonnenen Einzelspektren werden auf die gleiche Summe aller Spektralwerte

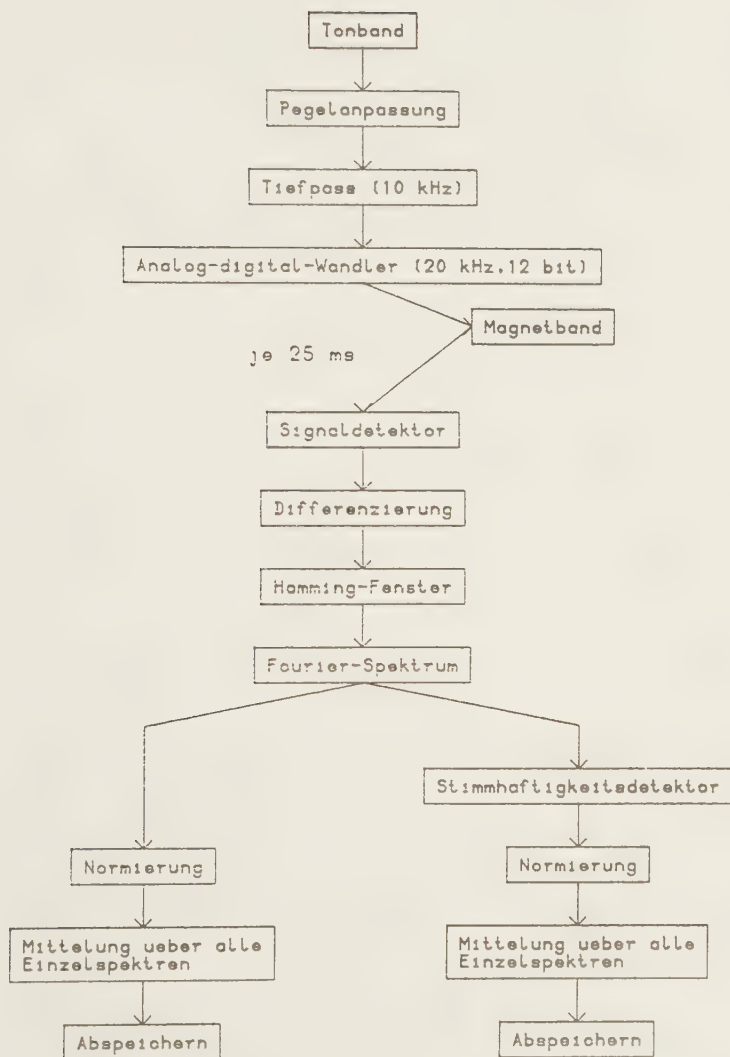


Abbildung 1:

Digitalisieren der Sprechproben und Berechnung der Langzeitspektren

normiert und aufsummiert. Das Summenspektrum wird dann durch die Anzahl der eingegangenen Einzelspektren dividiert und auf Magnetplatte abgespeichert.

Ein zweites Langzeitspektrum wurde von jeder Probe berechnet, in das bevorzugt stimmhafte Laute eingehen sollten. Hierfür wurde als Bedingung der Einzelspektren für die Aufnahme ins Langzeitspektrum verlangt, daß der Intensitätsanteil unterhalb 3000 Hz größer sein soll als der oberhalb 3000 Hz.

Zusammenfassend ist die Berechnung der Langzeitspektren in Abbildung 1 dargestellt.

Mit den Langzeitspektren wurden dann Paarvergleiche durchgeführt; dabei dienten als Abstands- bzw. Ähnlichkeitsmaß die Größe X^2 und der Korrelationskoeffizient r .

$$r = \frac{\sum_i [(A_i - \bar{A}) \cdot (B_i - \bar{B})]}{\sqrt{\sum_i (A_i - \bar{A})^2 \cdot \sum_i (B_i - \bar{B})^2}}$$

$$X^2 = \sum_i \frac{(A_i - B_i)^2}{A_i + B_i}$$

$\{A_1, A_2, \dots\} = \text{Spektrum A}, \quad \{B_1, B_2, \dots\} = \text{Spektrum B}$

3.1.2. Ergebnisse

Folgende Vergleichspaare können aus den 24 Langzeitspektren gebildet werden: Da von jedem der 10 Sprecher 2 Proben vorhanden sind, gibt es 10 intraindividuelle textverschiedene Vergleichspaare. Es kann auch jeder Sprecher mit jedem anderen verglichen werden, und zwar in jeder Kombination der zwei Textstücke. Daraus resultieren $4 \cdot 45 = 180$ interindividuelle Vergleichspaare. Diese setzen sich aus 90 textgleichen und 90 textverschiedenen zusammen. Abbildung 2 zeigt die Anordnung der Paare auf der Ähnlichkeitsachse für das Langzeitspektrum ohne Stimmhaftigkeitsbedingung und X^2 als Abstandsmaß.

Berechnet man von den intraindividuellen Paaren den Mittelwert und die Standardabweichung s , so zeigt sich, daß sämtliche interindividuellen Paare außerhalb des $3s$ -Bereichs oberhalb des Mittelwerts der intraindividuellen Paare liegen.

Auf die Sprecheridentifizierungs-Fragestellung angewendet, entspricht dies einer Erkennungsrate von 100% unter den gegebenen Versuchsbedingungen.

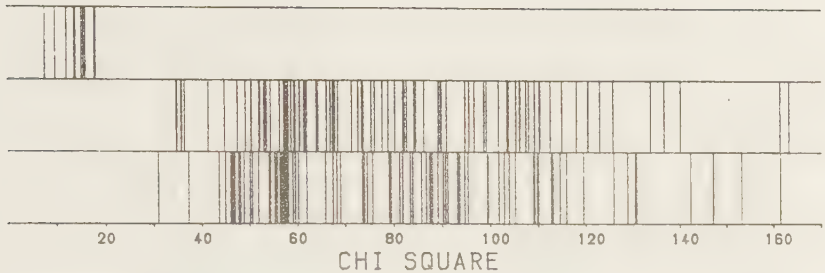


Abbildung 2:

Verteilung der intraindividuellen (oben) und der textgleichen (Mitte) und textverschiedenen (unten) interindividuellen Vergleichspaare (jeder senkrechte Strich stellt den Abstandswert eines Vergleichspaars dar).

Diese, verglichen mit den Studien von Majewski und Hollien (1975) und Zalewski, Majewski und Hollien (1975) höhere Erkennungsrate trotz kürzerer Integrationszeit bei der Erstellung der Langzeitspektren (die genannten Autoren arbeiteten mit 32 und 96 Sekunden) könnte daran liegen, daß die digital erstellten Langzeitspektren eine bessere Frequenzauflösung gegenüber den mit einem Terzfilter – Analysator erstellten haben, besonders in den höheren Frequenzbereichen.

Ein anderer Grund könnte allerdings auch in der Versuchsanordnung liegen, indem entweder die Sprecher – und Aufnahmesituation bei Majweski et al. nicht so konstant gehalten wurden oder indem die Sprechproben eines Sprechers zeitlich weiter auseinander lagen.

Kein signifikanter Unterschied besteht zwischen textgleichen und textverschiedenen Vergleichspaaren, wie aus Abbildung 2 hervorgeht. Danach kann man schließen, daß die Variabilität im Langzeitspektrum wesentlich stärker durch Sprecherunterschiede als durch Textunterschiede zustandekommt.

Dieselben Vergleiche wurden auch mit dem Korrelationskoeffizienten als Ähnlichkeitsmaß durchgeführt, was zu ähnlichen Ergebnissen führte. Ebenfalls zeigte sich eine vollständige Abtrennung der intraindividuellen Vergleichspaare von den interindividuellen; lediglich die Rangfolge zeigte geringfügige Unterschiede.

Bei Verwendung der Langzeitspektren, die mit dem 3000 – Hz – Stimmhaftigkeitsdetektor berechnet wurden, war die Kluft zwischen intra – und interindividuellen Paaren nicht ganz so groß, was heißt, daß durch die Unterdrückung hochfrequenter Signalstücke sprecherspezifische Information verloren geht.

Nun wurde noch eine weitere Art von Paarvergleichen durchgeführt:

Nach Berechnung des Differenzenspektrums wurden bei jedem Vergleichspaar die beiden Frequenzen notiert, bei denen der höchste bzw. niedrigste Wert im Differenzenspektrum zu finden war. Diese Werte lagen alle zwischen 200 und 4000 Hz. Vergleicht man jeweils dieselben Sprecher in verschiedenen Textstellen miteinander, so reproduzieren sich die Extrema entweder exakt oder sie liegen relativ weit auseinander, was man so interpretieren kann, daß es sich um sehr scharfe, in ihrer Lage stabile Extrema handelt, daß aber manchmal mehrere etwa gleich hohe Maxima oder Minima miteinander konkurrieren und einmal das eine, einmal das andere den höheren Wert aufweist.

Daß die Extrema sich, wenn überhaupt, exakt und nicht nur ungefähr reproduzieren, legt den Schluß nahe, daß in sehr eng begrenzten Frequenzbereichen, deren Breite 40 Hz nicht wesentlich übersteigt, das heißt in der Feinstruktur des Spektrums, Sprechereigenschaften zum Ausdruck kommen.

3.2. Kasuistische Beobachtung an zwei eineiigen Zwillingspaaren

Mit der in 3.1.1. beschriebenen Methode wurden auch ein männliches und ein weibliches eineiiges Zwillingspaar aufgenommen und in die Paarvergleiche miteinbezogen. Beim weiblichen Zwillingspaar besteht eine hohe Ähnlichkeit der Spektren; die χ^2 -Werte liegen genau in der in Abbildung 2 ersichtlichen Kluft zwischen intra- und interindividueller Variabilität.

Beim männlichen Zwillingspaar war im Langzeitspektrum kaum eine Ähnlichkeit zu beobachten; die χ^2 -Werte liegen sogar eher im oberen Bereich der interindividuellen Variabilität. Dies ist insofern erstaunlich, als die Stimmen der Zwillingsbrüder für das Ohr durchaus ähnlich klingen und daher am Telefon auch oft verwechselt werden.

Bezüglich ihrer Umwelthomogenität sind beide Zwillingspaare vergleichbar (gemeinsames Aufwachsen, dieselbe Schule, ähnliche Studiengänge der Geschwister). Aufnahmetechnische Gründe (Mikrophonanordnung, Raumakustik ...) sind zwar für die Erklärung der beobachteten Unähnlichkeit nicht ganz auszuschließen, dürften aber doch keine so gravierenden Unterschiede hervorrufen.

Zur Klärung dieses uneinheitlichen Ergebnisses müßte man eine größere Stichprobe von Zwillingspaaren untersuchen; im Rahmen dieser Arbeit wurde dies jedoch nicht weiterverfolgt, da die Interpretation von Zwillingsähnlichkeiten in der Stimme ähnlich problematisch ist wie bei der Intelligenzforschung (die Sprache wird größtenteils durch Nachahmung gelernt).

3.3. Erste Anhaltspunkte für einen Geschlechtsdimorphismus im Langzeitspektrum

Bei den Paarvergleichen aus 3.1. kann man die Paare auch in gleichgeschlechtliche und verschiedengeschlechtliche unterteilen und sich die Frage stellen, ob das Geschlecht bei den Ähnlichkeitsvergleichen zum Ausdruck kommt.

Stellt man sämtliche interindividuellen X^2 -Werte in einem 20-Klassen-Häufigkeitsdiagramm dar, so zeigt sich eine leichte Tendenz zur Zweigipfeligkeit (Abbildung 3), die sich weitgehend auflöst, wenn man gleich- und verschiedengeschlechtliche Paare getrennt aufzeichnet (Abbildung 4). Der Abstand der beiden Maxima zeigt, in welcher Größenordnung Sprecherunterschiede liegen, die dem Geschlechtsdimorphismus zugeschrieben werden können.

Die Auflösung der Zweigipfeligkeit ist beiden Langzeitspektren mit Stimmhaftigkeitsdetektor etwas ausgeprägter als bei den normalen.

Als ein weiterer Schritt zu Charakterisierung des Geschlechtsdimorphismus wurde versucht, einfache Parameter im Langzeitspektrum zu finden, die sich als dimorph erweisen.

Hierfür wurden die Spektren durch senkrechte Trennlinien in vier flächengleiche Teile zerlegt und die Frequenzwerte der Trennlinien (Quartilgrenzen) notiert. Als erster Parameter wurde die obere Grenze des untersten Quartils verwendet; wie Abbildung 5 oben links zeigt, liegt sie bei Männern im Schnitt höher als bei Frauen. Das heißt, daß der relative Intensitätsanteil in den unteren Frequenzbändern bei Männern niedriger ist als bei Frauen. Dies scheint zunächst paradox, da man ja weiß, daß Männer im Schnitt tiefere Stimmen haben. Da der erste Quartil aber mindestens bis 800 Hz reicht, sind in ihm noch einige Harmonische der Grundfrequenz enthalten; außerdem macht die Grundfrequenz selbst nur einen geringen Teil der Gesamtintensität des Langzeitspektrums aus.

Wenn tatsächlich ein Geschlechtsdimorphismus in der Intensität der unteren Harmonischen der Grundfrequenz besteht, müßte dieser bei den Langzeitspektren, die mit dem 3000 Hz-Stimmhaftigkeitsdetektor erstellt wurden, stärker zum Ausdruck kommen, was auch aus Abbildung 5 unten ersichtlich ist.

Um einen weiteren möglichst unabhängigen Parameter zu bekommen, wurde der prozentuale Anteil der Breite des vierten Quartils an der Gesamtbreite der Quartile 2, 3 und 4 berechnet (Abbildung 5 rechts). In diesem Parameter weisen die männlichen Sprecher im Schnitt höhere Werte auf als die Frauen.

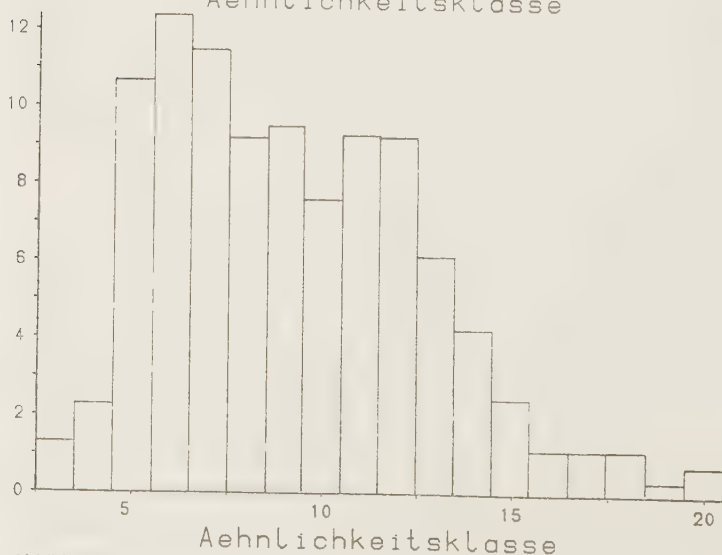
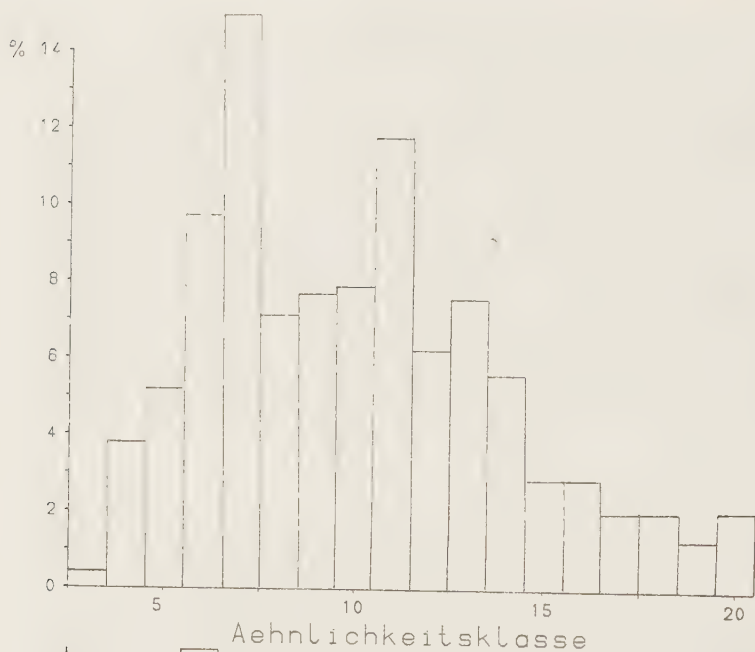


Abbildung 3:

Verteilung der interindividuellen χ^2 -Werte

oben: nach normalem Langzeitspektrum

unten: nach Langzeitspektrum mit Stimmhaftigkeitsdetektor

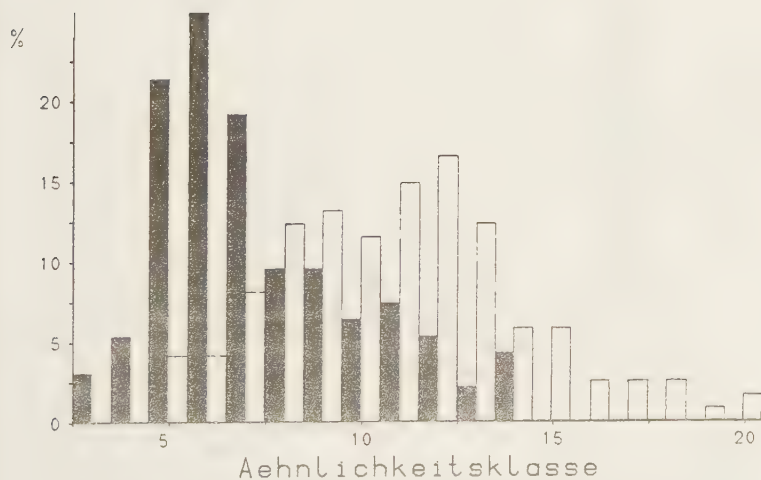
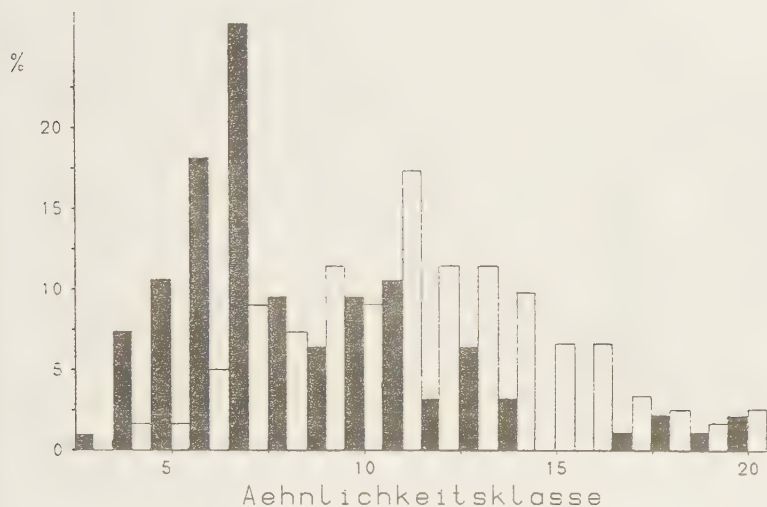


Abbildung 4:

Auftrennung der in Abbildung 3 dargestellten χ^2 -Werte in Werte gleichgeschlechtlicher (weiß) und verschiedengeschlechtlicher (schwarz) Vergleichspaare:

Die zu vermutende Zweigipfeligkeit in Abbildung 3 löst sich dadurch weitgehend auf. Bei Langzeitspektren mit Stimmhaftigkeitsdetektor (unten) ist diese Auflösung noch etwas ausgeprägter.

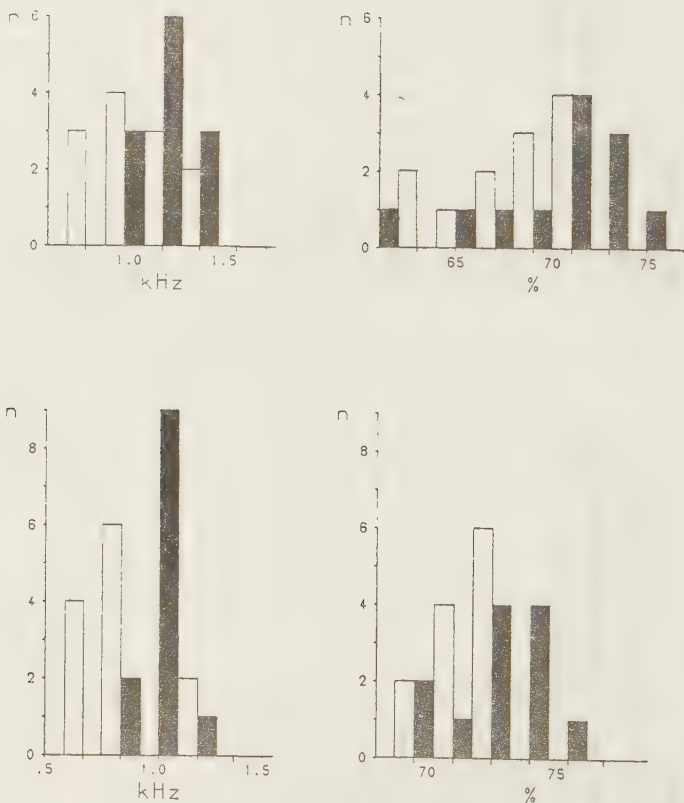


Abbildung 5:

Geschlechtsdimorphismus in der Lage der Quartile des Langzeitspektrums:

oben: normales Langzeitspektrum

unten: Langzeitspektrum mit Stimmhaftigkeitsdetektor

links: Verteilung der oberen Grenze des ersten Quartils

rechts: Anteil des 4. Quartils am Restspektrum (ohne 1. Quartil)

(Siehe auch Text Seite 31)

3.4. Beobachtung bezüglich Sprachunabhängigkeit des Langzeitspektrums

Mit der in 3.1.1. beschriebenen Methode wurden 4 Langzeitspektren einer Sprecherin gemacht, die Polnisch als Muttersprache spricht. Die Tonaufnahmen wurden für diesen Zweck vom Institut für Phonetik und sprachliche Kommunikation zur Verfügung gestellt. Zwei der Proben sprach die Sprecherin auf Deutsch und zwei auf Polnisch.

Der Paarvergleich der Langzeitspektren mit χ^2 als Ähnlichkeitsmaß brachte folgendes Ergebnis: Am ähnlichsten waren sich jeweils die gleichsprachigen Proben; die verschiedensprachigen lagen aber nur leicht höher, ungefähr in dem Bereich, in dem die Geschwisterwerte des in 3.2. beschriebenen weiblichen Zwillingspaars lokalisiert sind; sie setzen sich noch deutlich von den interindividuellen Werten ab.

Obwohl es sich hier nur um eine Einzelbeobachtung handelt, kann man doch vermuten, daß die Sprache im Langzeitspektrum weniger zum Ausdruck kommt als die Individualität des Sprechers.

3.5. Einfluß von Sprechtonhöhe und Lautstärke auf das Langzeitspektrum

Diese Untersuchung wurde mit dem Ziel durchgeführt, Parameter zu finden, die stabil sind gegenüber einfachen willentlichen Einflüssen von Seiten des Sprechers.

Auch einem ungeübten Sprecher bereitet es keine Schwierigkeiten, seine Lautstärke und seine Tonhöhe zu verändern; daher wurden diese beiden Variationsmöglichkeiten hier ausgewählt.

3.5.1. Material und Methoden

Es wurden Tonaufnahmen von 4 Versuchspersonen erstellt:

- (A) männlich, ca. 30 Jahre
- (B) weiblich, ca. 60 Jahre
- (C) weiblich, ca. 60 Jahre
- (D) weiblich, ca. 30 Jahre

(C) und (D) haben auditiv sehr ähnliche Stimmen (es handelt sich um Mutter und Tochter). Die Sprecher bekamen die Anweisung, denselben Text je einmal leise, normal, laut, tief und hoch (jeweils in dieser Reihenfolge) zu sprechen. Von Sprecher (A) wurden zur Kontrolle zusätzlich 6 Aufnahmen mit verschiedenen Mikrofonstellungen und in

unterschiedlichen Räumen gemacht. Als Text wurde eine Passage von ca. 12 Sekunden aus "der Nordwind und die Sonne" verwendet.

Die Aufnahmen wurden mit einem Tonbandgerät Uher Report Monitor 4200 und dem dazugehörigen Mikrophon Uher 518a (leichte Richtcharakteristik) erstellt; Bandlaufgeschwindigkeit 19 cm/s; Mund – Mikrophon – Abstand ca. 30 cm.

Da die Verarbeitung der Sprechproben am Institut für Phonetik und sprachliche Kommunikation aus Kapazitätsgründen nicht uneingeschränkt möglich war, wurde für diese und weitere Untersuchungen am Leibnitz – Rechenzentrum der Bayerischen Akademie der Wissenschaften ein neues Auswertungsprogramm erstellt. Dies hat sich auch insofern als vorteilhaft erwiesen, als dort auch die Möglichkeiten für die statistische Bearbeitung der Ergebnisdaten gegeben sind.

Da das Rechenzentrum aber über keinen analog – digital – Wandler verfügt, mußten die Sprechproben zuvor digitalisiert und auf Magnetbänder übertragen werden. Dies geschah am Lehrstuhl für Datenverarbeitung der TU München:

Die Proben wurden auf demselben Tonbandgerät abgespielt, auf dem sie aufgenommen wurden (um zusätzliche Geräteartefakte zu vermeiden), Tiefpaß – gefiltert (8000 Hz, 48 db/Oktave), digitalisiert (20000 Hz, 12 bit) und auf einem lokalen Plattenspeicher abgelegt. Von dort konnten sie noch einmal zur Kontrolle über Lautsprecher abgehört werden, bevor sie auf Magnetband abgespeichert wurden (siehe Abbildung 6). Das Rechenprogramm zur Erstellung der Langzeitspektren wurde in FORTRAN 77 programmiert und entspricht weitgehend dem in Abbildung 1 abgebildeten Schema; ein Unterprogramm für die schnelle Fourier – Transformation war am Rechenzentrum vorhanden. Der 3000 – Hz – Stimmhaftigkeitsdetektor wurde weggelassen, da er einerseits nicht wesentlich dazu beiträgt, die interindividuelle Variabilität im Sprachsignal herauszuarbeiten, andererseits aber die Menge der zu speichernden Ergebnisdaten verdoppelt.

Die 256 Werte des Langzeitspektrums sind allerdings immer noch eine große Datenfülle. Um sie zu reduzieren, gibt es die Möglichkeit, nach einer Glättung der Spektren nur jeden zweiten Wert zu verwenden. Dadurch geht zwar Information verloren; da aber Nachbarwerte relativ stark miteinander korreliert sind, dürfte dieser Verlust nicht allzu groß sein. Für die Glättung wurde ein 18 – stufiges nicht – rekursives digitales Tiefpaßfilter verwendet, das sehr steil bei einem Viertel der Sampling – Frequenz schneidet. Die optische Kontrolle zeigt, daß sich durch dieses Filter die Spektren nur unwesentlich verändern (siehe Abbildung 7). Der Bereich oberhalb 8000 Hz wurde weggelassen, so daß sich die zu speichernden Spektralwerte auf 100 reduzieren.

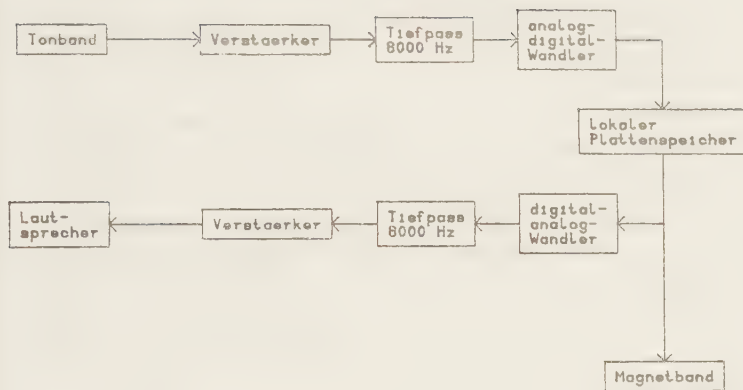


Abbildung 6:

Digitalisieren der Sprechproben

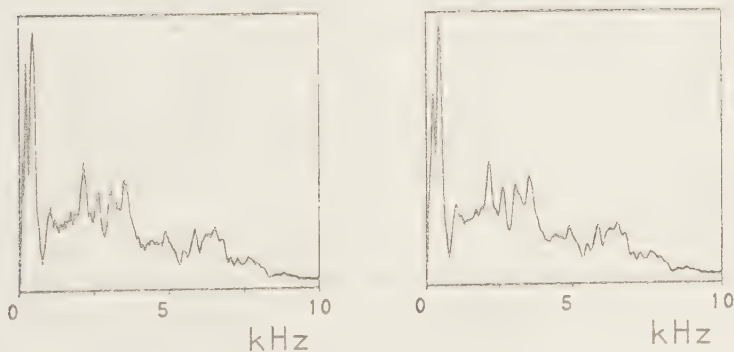


Abbildung 7:

links: Langzeitspektrum mit 256 40-Hz-Bändern; rechts: Dasselbe Spektrum nach Glättung mit einem 18-stufigen digitalen Tiefpaßfilter

Als nächster Schritt wurde versucht, aus dem Langzeitspektrum sekundäre Parameter zu gewinnen, die folgenden Anforderungen genügen:

- (1) Anschaulichkeit bezüglich der Frequenzskala,
- (2) hohes Maß an Sprecherspezifität,
- (3) geringe Datenmenge.

Hierfür wurden die Spektren in folgende Teilstrukturen zerlegt: Grobverlauf, mittelgrober Verlauf, mittelfeiner Verlauf und Feinstruktur. Dies wurde realisiert, indem das Spektrum mit Hilfe der Fourier – Transformation in Sinus – und Cosinuskomponenten zerlegt und nur ein Teil der Komponenten jeweils wieder rücktransformiert wurde (siehe Abbildung 8). Als Schnittstellen wurden Cepstralbereiche genommen, die im Mittel (über alle Sprechproben) eine geringe Amplitude aufweisen.

3.5.2. Ergebnisse

Für einen optischen Vergleich der strukturzerlegten Spektren wurden zunächst Graphiken auf transparentes Papier gezeichnet und durch Übereinanderlegen verglichen. Dann wurden verschiedene Kombinationen zur Dokumentation des Gesehenen übereinandergezeichnet.

Abbildung 9 zeigt die intraindividuelle Variabilität: Von jedem Sprecher sind die leise, laut, tief und hoch gesprochenen Proben gemeinsam abgebildet.

In Abbildung 10 sind jeweils die vier Sprecher übereinandergezeichnet; dadurch soll die interindividuelle Variabilität bei gleicher Sprechintention gezeigt werden.

Sucht man nun sprecherstabile Merkmale, in denen sich verschiedene Sprecher unterscheiden, so muß man fordern, daß die Streubreite dieser Merkmale in Abbildung 9 gering und in Abbildung 10 hoch ist. Diese Bedingung erfüllen vor allem die mittelgrobe und die mittelfeine Struktur im Bereich zwischen 1000 und 5000 Hz sowie die Feinstruktur im Bereich zwischen 1000 und 3000 Hz.

Die Grobstruktur erwies sich in erster Linie als Lautstärke – abhängig (siehe auch Abbildung 8); bei laut gesprochenen Proben war bei allen Sprechern die Intensität zwischen ca. 2000 und 3000 Hz deutlich angehoben.

Inwieweit die Teilstrukturen auch stabil sind gegenüber Mikrophonstellung und leichten Unterschieden in der Raumakustik, zeigt Abbildung 11 oben links: Dort sind 6 Spektren von Sprecher (A) übereinandergezeichnet; die zugrundeliegenden Tonaufnahmen wurden in drei verschiedenen Räumen (Küche, Arbeitszimmer und Wohnzimmer) und mit drei verschiedenen Mikrophonstellungen (Variation des Aufnahmewinkels zur Sprechrichtung



Abbildung 8:

Vier Langzeitspektren mit unterschiedlicher Sprechweise eines Sprechers, die in Teilstrukturen zerlegt sind:

oben links: leise; oben rechts: laut;

unten links: tief; unten rechts: hoch.

Die Teilstrukturen sind (von oben nach unten):

Fein-, Mittelfein-, Mittelgrob- und Grobstruktur (gestrichelt).

Das Originalspektrum ist jeweils die untere durchgezogene Linie.



Abbildung 9:

Intraindividuelle Variabilität im strukturzerlegten Langzeitspektrum bei unterschiedlicher Sprechweise:

Von Sprecher (A) (oben links), (B) (oben rechts), (C) (unten links) und (D) (unten rechts) sind jeweils vier Spektren mit unterschiedlicher Sprechweise (leise, laut, tief, hoch) übereinandergezeichnet.

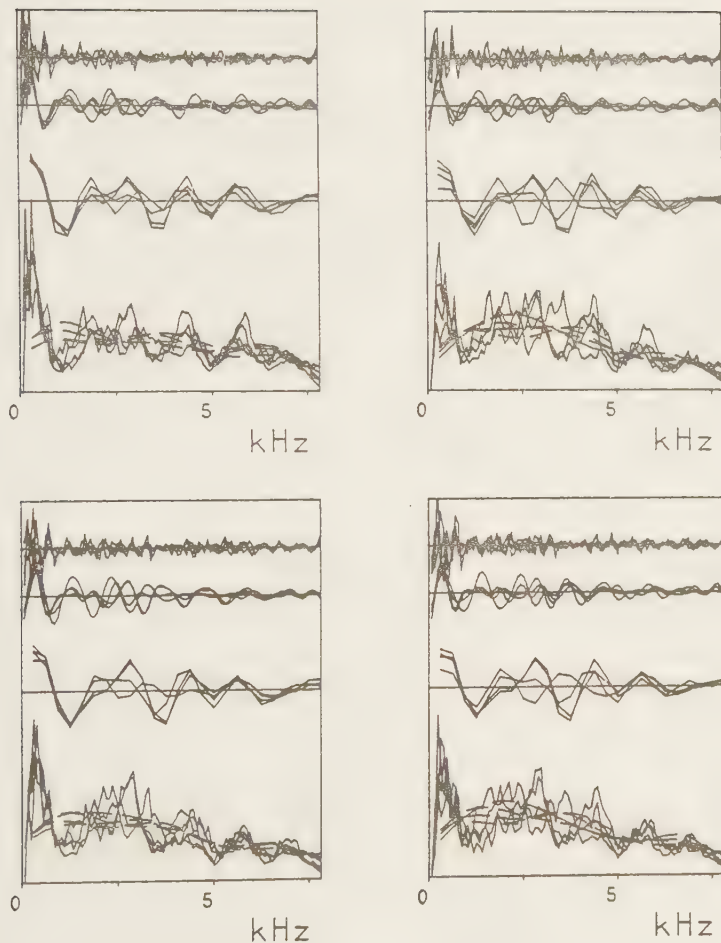


Abbildung 10:

Interindividuelle Variabilität in strukturzerlegten Langzeitspektrum:

Die vier Sprecher (A), (B), (C) und (D) sind jeweils bei gleicher Sprechweise übereinandergezeichnet:

oben links: leise; oben rechts: laut;

unten links: tief; unten rechts: hoch.

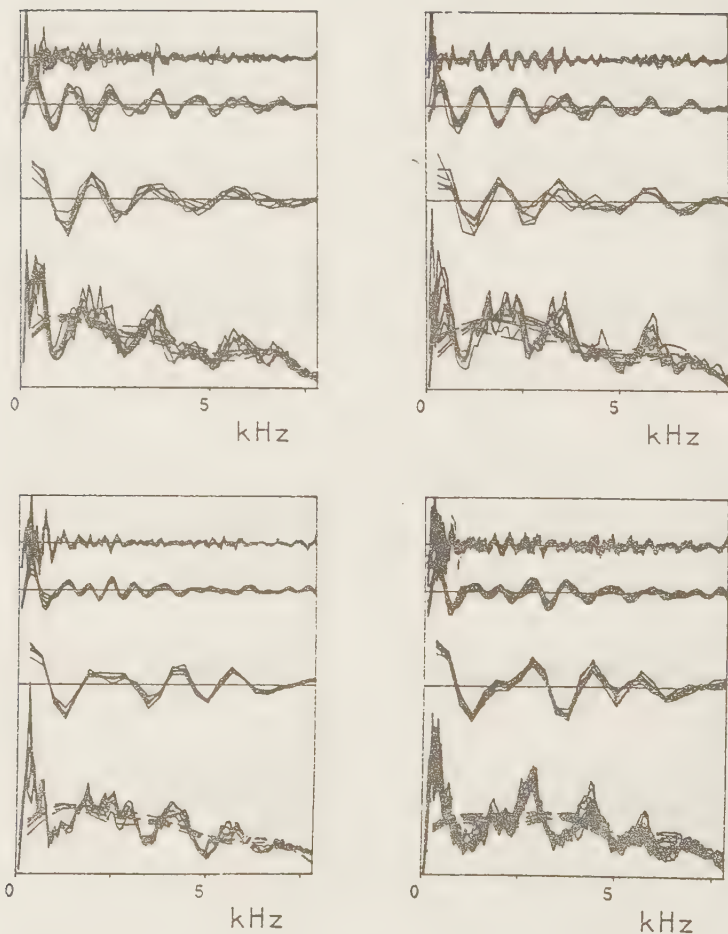


Abbildung 11:

Vergleich der Streuung im strukturzerlegten Langzeitspektrum durch Aufnahmebedingungen (oben links, Sprecher (A) in drei verschiedenen Räumen und mit unterschiedlichen Mikrophonstellungen), Sprechweise (oben rechts und unten links, Sprecher (A) und (B), jeweils leise, normal, laut, tief und hoch) und durch Sprechweise, wenn man zwei sehr ähnliche Sprecher gemeinsam abbildet (unten rechts, Sprecher (C) und (D), beide leise, normal, laut, tief und hoch).

von 30 bis 60 Grad) durchgeführt.

Im Gegensatz zu Abbildung 11 oben rechts (Sprecher (A) leise, normal, laut, tief und hoch) zeigt sich hier die Feinstruktur wenig stabil; sie weist eine mit Abbildung 10 vergleichbare Streuung auf.

Daraus kann man schließen, daß die Feinstruktur ungefähr in derselben Größenordnung von den technischen Gegebenheiten abhängt, in der auch die interindividuellen Differenzen liegen. Daher wurde sie für die weiteren Untersuchungen nicht mehr verwendet. (Bei einer Serie von Aufnahmen, die im Laufe mehrerer Monate gemacht werden, ist es schwierig, die Aufnahmebedingungen exakt konstant zu halten.)

Eventuell enthält die Feinstruktur monofrequente, für das Ohr unterschwellige Hintergrundgeräusche, die sich im Langzeitspektrum akkumulieren und kleine, spitze Peaks hervorrufen.

Bezüglich Grobverlauf zeigt Abbildung 11 oben links eine geringere Streubreite als Abbildung 11 oben rechts; das heißt, daß der Grobverlauf mehr von der Sprechweise als von der Aufnahmeanordnung abhängt.

Es wurde bereits erwähnt, daß die Versuchspersonen (C) und (D) (Mutter und Tochter) sehr ähnliche Stimmen haben. Wie sich diese Ähnlichkeit im Langzeitspektrum ausdrückt, zeigt Abbildung 11 unten rechts. Dort sind alle 10 Proben der beiden Sprecherinnen übereinandergezeichnet; die beobachtbare Streuung liegt etwa zwischen der intraindividuellen Streuung von (A) (oben rechts) und (B) (unten links). Da der Altersunterschied von (C) und (D) ca. 30 Jahre beträgt, kann aus dieser Einzelbeobachtung heraus vermutet werden, daß sich das Lebensalter von Erwachsenen im Langzeitspektrum (zumindest sprecherunabhängig) kaum ausdrückt.

Ein Problem besteht noch darin, daß mit den hier beschriebenen Methoden der Frequenzbereich unterhalb 1000 Hz nicht mit hinreichender Genauigkeit erfaßt werden kann. Insbesondere Unterschiede in der Sprechtonhöhe gehen bei einer Frequenzauflösung von 80 Hz verloren.

Daher sei dieser Frequenzbereich hier auch nicht näher betrachtet; in Kapitel 3.8. wird er gesondert untersucht.

Als ein weiterer Schritt zur Untersuchung der inter- und intraindividuellen Variabilität wurde nun untersucht, inwieweit die vier Sprecher mit statistischen Methoden auseinandergehalten werden können.

Zunächst wurde jede Probe mit jeder im gesamten Langzeitspektrum mit Hilfe des

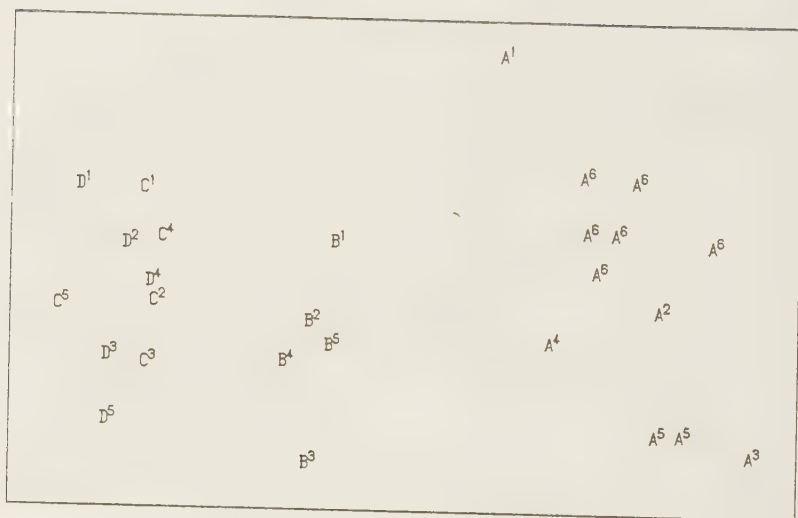


Abbildung 12:

Zweidimensionale Darstellung von Sprecherähnlichkeiten im Langzeitspektrum mit Hilfe der "smallest space analysis" (siehe auch Text Seite 45);

Buchstaben: Sprecher (A), (B), (C) und (D);

Zahlen: 1 = leise, 2 = normal, 3 = laut, 4 = tief, 5 = hoch,

6 = normal bei unterschiedlichen Aufnahmebedingungen

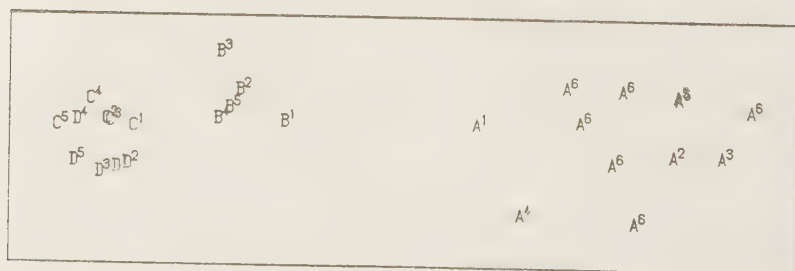


Abbildung 13:

Zweidimensionale Darstellung von Sprecherähnlichkeiten in der mittleren Struktur des Langzeitspektrums im Bereich von 1000 bis 5000 Hz (Symbole analog Abbildung 12)

Korrelationskoeffizienten verglichen. Zur graphischen Darstellung der resultierenden Ähnlichkeitsmatrix wurde die "smallest space analysis" eingesetzt: Die Proben werden in einem Raum möglichst weniger Dimensionen so angeordnet, daß ähnliche Proben näher beieinander liegen als weniger ähnliche. Zur Vereinfachung und besseren graphischen Darstellbarkeit wurde die Zahl der Dimensionen immer auf zwei begrenzt (unter Vernachlässigung der Restvarianz). Die Berechnung der Koordinaten geschah mit Hilfe des Unterprogramms "SSA1" aus dem SPSS- Statistik-Programmpaket (Toledano 1977). Den Koordinaten selbst kommt nur eine relative Bedeutung zu; sie reichen in jeder Dimension von 0 bis 100; daher wurde in den Abbildungen 12 und 13, die Ausschnitte aus solchen Space-Diagrammen darstellen, auf eine Achsenbeschriftung verzichtet.

Abbildung 12 zeigt das Ähnlichkeitsgefüge der Sprechproben. Die Buchstaben sind die Sprecher, die Zahlen 1 bis 5 bedeuten die Sprechweise, 6 bedeutet normal gesprochene Aufnahme mit veränderter Mikrophonstellung und Raumakustik.

Bis auf (C) und (D) zeigt sich eine deutliche Sprechertrennung; die Proben mit unterschiedlichen Aufnahmebedingungen von (A) sind sich ähnlicher als die mit unterschiedlicher Sprechweise desselben Sprechers.

Aufgrund der Beobachtungen, daß die mittelgrobe und die mittelfine Struktur relativ sprecherkonstant sind, wurden nun diese beiden Strukturen zur mittleren Struktur zusammengefaßt. Diese besitzt eine Frequenzauflösung von 240 bis 1000 Hz (bei 1000 Hz fängt die Frequenzauflösung der Grobstruktur an) und kann im Bereich von 1000 bis 5000 Hz durch 17 Einzelwerte mit hinreichender Genauigkeit beschrieben werden.

Diese 17 Werte wurden für jede Sprechprobe auf denselben mittleren Absolutbetrag normiert und abgespeichert. Nun wurde die so erfaßte mittlere Struktur des Langzeitspektrums als Kriterium für Paarvergleiche verwendet.

Abbildung 13 zeigt das Ergebnis analog Abbildung 12: Die Sprechertrennung ist gegenüber Abbildung 12 etwas verbessert; insbesondere zeigt sich bei (C) und (D) eine Tendenz zu Gruppierung, lediglich die mit hoher Stimme von (C) gesprochene Probe kann mit der mit tiefer Stimme von (D) gesprochenen verwechselt werden. Bei (A) rückt die durch die Sprechweise bedingte Streuung in den Bereich, von der durch die Aufnahmebedingungen verursachten (A^6) Streuung, vermutlich deshalb, weil die unterschiedliche Sprechweise hauptsächlich in der hier eliminierten Grobstruktur des Spektrums zum Ausdruck kommt.

Eine weitere Methode, Gruppenunterschiede nachzuweisen, ist die Diskriminanzanalyse. Als Gruppe seien hier jeweils die Proben eines Sprechers bezeichnet; es gibt also vier

Gruppen. Da die Diskriminanzanalyse nur angewendet werden kann, wenn die Zahl der eingehenden Variablen klein ist gemessen an der Zahl der Proben, wurden nur die ersten 7 Werte der mittleren Struktur verwendet. Diese 7 Werte beinhalten den Spektralverlauf der mittleren Struktur des Langzeitspektrums zwischen ca. 1000 und 2500 Hz, mit einer Frequenzauflösung von 240 Hz.

Da es schwierig ist, bei kleinen Gruppengrößen Signifikanztests bei der Diskriminanzanalyse zu interpretieren, wurde die Überprüfung der Gruppentrennung empirisch vorgenommen: Es wurden fünf verschiedene Diskriminanzanalysen gerechnet; bei jeder wurde je eine Probe von jedem Sprecher in der Analysephase weggelassen und dann automatisch der wahrscheinlichsten Gruppe zugeordnet. Insgesamt wurden also 20 Zuordnungen vorgenommen.

In 19 Fällen wurde die Probe dem tatsächlichen Sprecher richtig zugeordnet; lediglich die tief gesprochene Probe von (D) wurde als (C) bezeichnet, was sich auch mit den in Abbildung 13 dargestellten Ähnlichkeitsbeziehungen vereinbaren läßt.

Zusammenfassend lassen sich aus dieser Untersuchung folgende Schlüsse ziehen:

- (1) Das Langzeitspektrum ändert sich nicht wesentlich, wenn ein Sprecher Lautstärke und Tonhöhe variiert.
- (2) Die Lautstärke drückt sich vor allem durch eine Anhebung der Intensität im Bereich von 2000 bis 3000 Hz aus.
- (3) Die Feinstruktur des Langzeitspektrums ist zwar stabil gegenüber Lautstärke und Tonhöhe, verändert sich aber empfindlich mit den Aufnahmebedingungen.
- (4) Die mittleren Strukturen des Langzeitspektrums sind sowohl relativ sprecherkonstant als auch nicht zu stark von der Aufnahmesituation abhängig.
- (5) Bereits mit 7 Parametern aus der mittleren Struktur des Langzeitspektrums ist mit Hilfe der Diskriminanzanalyse eine gute Sprechererkennung möglich.

3.6. Abhängigkeit des Langzeitspektrums von Mikrofon – und Raumakustik

Die Sprechproben und Langzeitspektren für diese Studie wurden vom Institut für Phonetik und sprachliche Kommunikation zur Verfügung gestellt.

Von zwei Sprechern wurden 13 Tonaufnahmen in drei verschiedenen Räumen und mit zwei verschiedenen Mikrofonen erstellt. Eines der Mikrophone hatte zusätzlich einen Sprache/Musik – Schalter. Die Spektren lagen auf Magnetband vor und konnten im Rahmen dieser Arbeit am Rechenzentrum eingelesen und statistisch ausgewertet werden. Sowohl die gesamten Spektren als auch die mit dem in 3.5.1. beschriebenen Verfahren berechnete mittlere Struktur (17 Parameter zwischen 1000 und 5000 Hz) wurden für die Ähnlichkeitsvergleiche verwendet.

Abbildung 14 zeigt die Anordnung der Proben nach Ähnlichkeiten in der mittleren Struktur des Langzeitspektrums (fast dasselbe Bild ergab die Verwendung der gesamten Spektren): Die Sprecher (F und G) werden in der ersten Dimension (von links nach rechts) vollständig getrennt. In der zweiten Dimension zeigt sich der Effekt eines Aufnahmerraums: alle Proben oberhalb der gestrichelten Linie wurden im Studio aufgenommen, die unterhalb im Hörsaal und im Sprachlabor. Bezüglich dieser beiden Räume und bezüglich Mikrofon sowie Sprache/Musik – Schalter zeigt sich keine Gruppierung.

Dieser Vortest wurde insbesondere im Hinblick auf die folgende Studie mit Sprecher – Imitationsversuchen gemacht, da die dort verwendeten Tonaufnahmen unter sehr verschiedenen Bedingungen zustandekamen.

In Abbildung 14 ist zwar die interindividuelle Streuung größer als die durch Apparate bedingte; bei Sprechern mit sehr ähnlichen Stimmen dürften jedoch unter so unterschiedlichen Aufnahmebedingungen Überschneidungen zu erwarten sein.

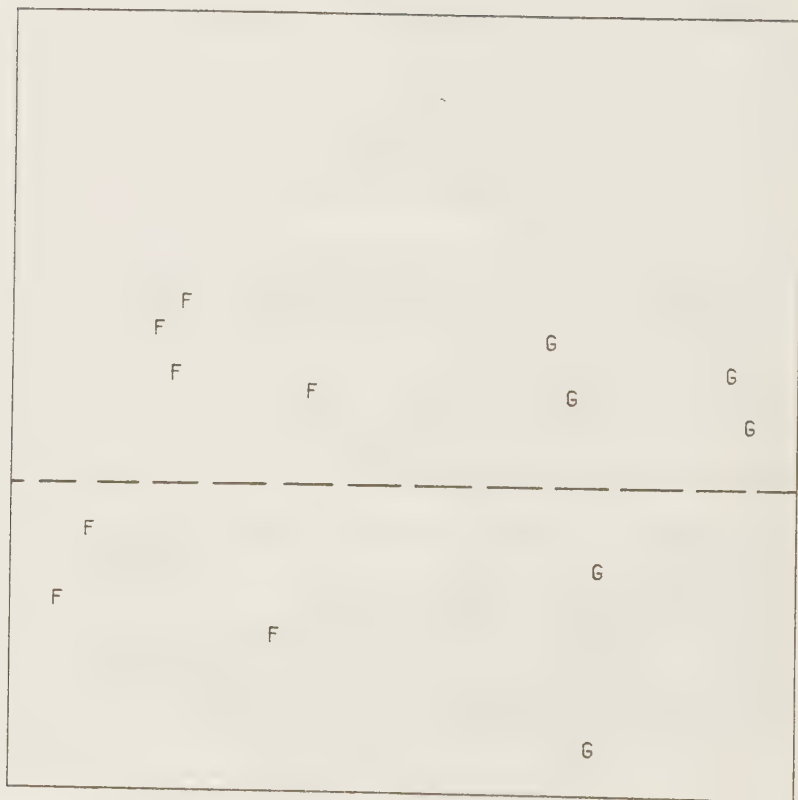


Abbildung 14:

Zweidimensionale Darstellung der Ähnlichkeiten zweier Sprecher (F und G) in der mittleren Struktur des Langzeitspektrums unter verschiedenen Aufnahmebedingungen (drei verschiedene Räume und zwei verschiedene Mikrophone):

In der ersten Dimension (rechts/links) findet die Auftrennung nach Sprechern statt; die zweite Dimension trennt alle in einem Raum erstellten Aufnahmen von den übrigen (durch die gestrichelte Linie gekennzeichnet). Effekte der unterschiedlichen Mikrophone zeichnen sich nicht ab.

3.7. Willentliche Beeinflußbarkeit des Langzeitspektrums bei Sprecher-Imitationsversuchen

Für diese Untersuchung wurden die Langzeitspektren der Sprechproben vom Institut für Phonetik und sprachliche Kommunikation zur Verfügung gestellt. Zweck dieser Studie ist, zu sehen, inwieweit ein geübter Sprecher in der Lage ist, seine Stimme gezielt zu verändern, inwieweit er noch identifizierbar ist und wie nah er in objektiven akustischen Parametern an den Imitierten herankommt.

Als anthropologischer Aspekt kann hier eine obere Grenze der intraindividuellen Variabilität in der Sprechstimme abgeschätzt werden.

Das Tonbandmaterial setzt sich folgendermaßen zusammen: Sprecher (A) ist ein Kabarettist, der in der Lage ist, Politikerstimmen so nachzuahmen, daß die Imitate für das menschliche Ohr kaum von den Originalen zu unterscheiden sind. Sprecher (B), (C), (D) und (E) sind bekannte Politiker; die Aufnahmen stammen größtenteils aus dem Bundestag. Die Imitationen wurden teilweise im Studio aufgenommen, teilweise von einer Schallplatte abgespielt.

Insgesamt wurden 29 Proben ausgewertet: 4 Proben (A) original, je 3 Proben von (B), (C), (D) und (E) original, 3 Imitationen (A/b), 2 Imitationen (A/c), 4 Imitationen (A/d) und 2 Imitationen (A/e). (Die Imitierten sind immer mit Kleinbuchstaben gekennzeichnet, die Sprecher mit Großbuchstaben)

Die Spektren lagen auf Magnetband vor und wurden am Rechenzentrum in Teilstrukturen zerlegt und statistisch ausgewertet.

Aus Abbildung 15 geht hervor, daß die intraindividuelle Variabilität bei den Imitationen (rechts oben) deutlich vergrößert ist gegenüber den Normalstimmen

Jede Probe wurde nun unter Verwendung des Korrelationskoeffizienten als Ähnlichkeitsmaß mit jeder im gesamten Langzeitspektrum sowie in der mittleren Struktur des Langzeitspektrums verglichen; die Korrelationsmatrizen wurden mit Hilfe der smallest space analysis als zweidimensionale Graphiken dargestellt (Abbildung 16 und 17).

Mit Ausnahme von Sprecher (E) sind in beiden Ähnlichkeitsräumen alle Originalstimmen gut nach Sprechern gruppiert; die mittlere Struktur des Langzeitspektrums (Abbildung 17) erscheint zur Trennung ebenso gut geeignet wie das Langzeitspektrum selbst (Abbildung 16) (obwohl sie nur 17 Parameter enthält gegenüber 100 Werten im geglätteten Langzeitspektrum).

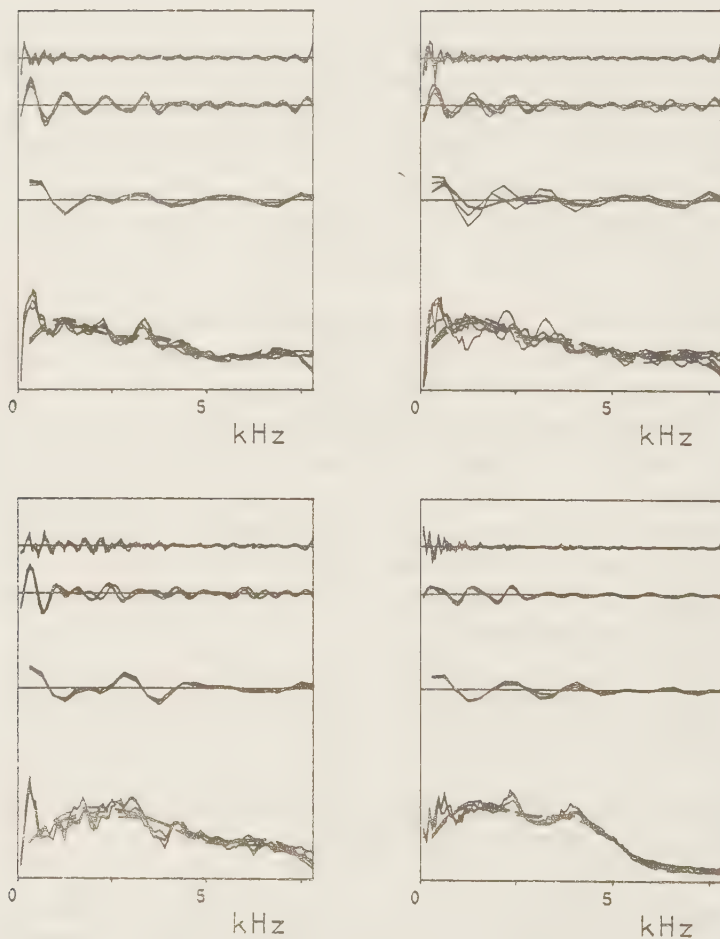


Abbildung 15:

Intraindividuelle Variabilität im strukturzerlegten Langzeitspektrum

oben links: normale Stimme des Imitators (4 Proben)

oben rechts: Stimme des Imitators bei vier verschiedenen Imitationen

unten links: drei Spektren des Sprechers (D)

unten rechts: drei Spektren des Sprechers (B)

Man sieht, daß sich die Variabilität beim Imitieren stark vergrößert.

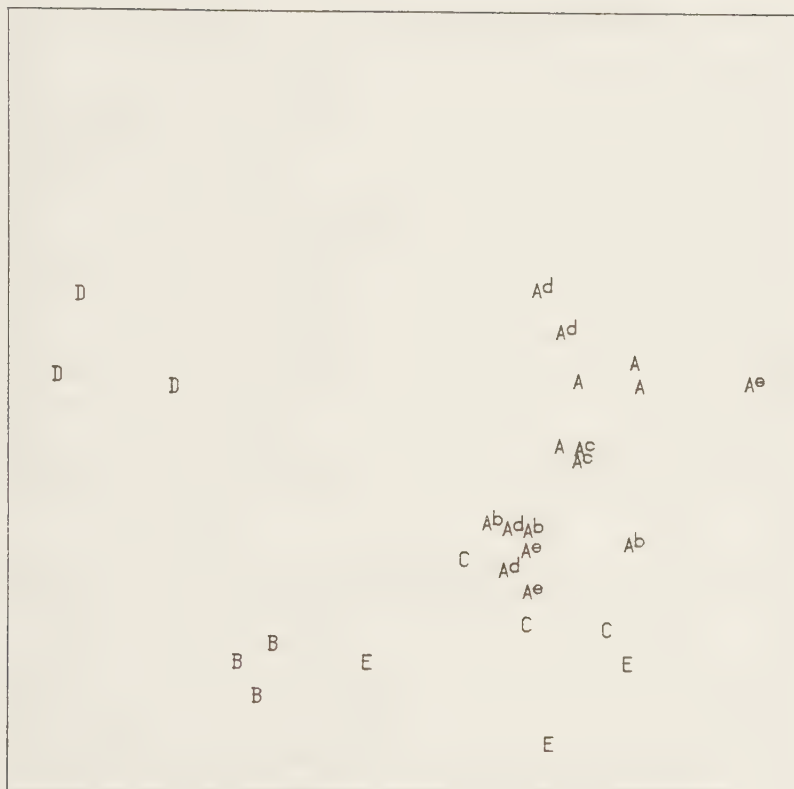


Abbildung 16:

Zweidimensionale Darstellung der Ähnlichkeiten von Sprechproben aus Imitationsversuchen (A ist der Imitator) im Langzeitspektrum:

Großbuchstaben: Sprecher Kleinbuchstaben: Imitierter

A^e bedeutet, daß Sprecher (A) versucht hat, Sprecher (E) nachzuahmen. Der Darstellung liegt das Ähnlichkeitsgefüge der Sprechproben im gesamten Langzeitspektrum zugrunde; als Ähnlichkeitsmaß wurde der Korrelationskoeffizient verwendet.

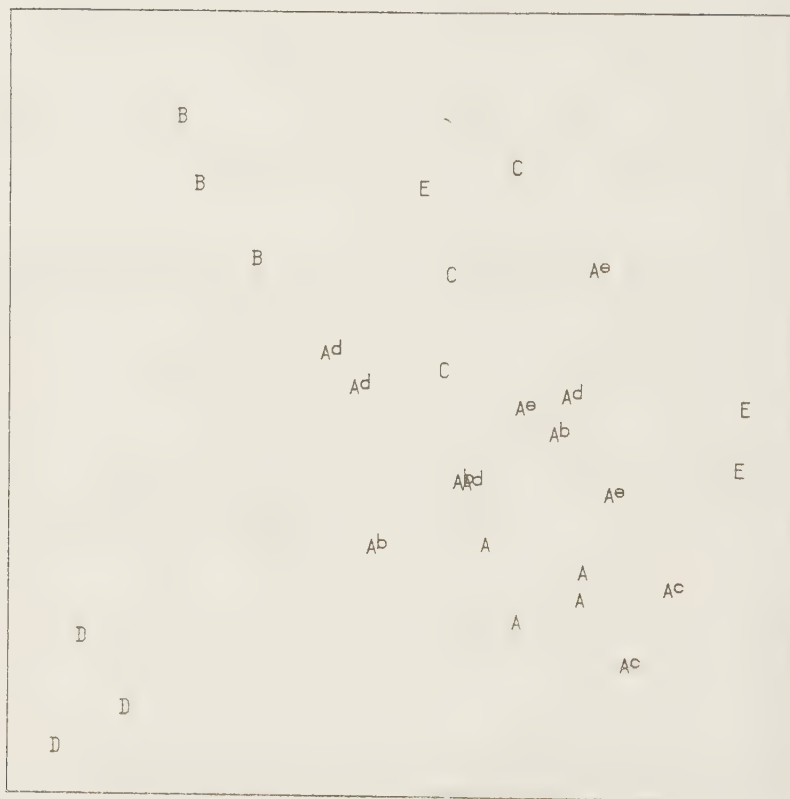


Abbildung 17:

Ähnlichkeit von Sprechproben bei Imitationsversuchen:

Der Darstellung liegen Ähnlichkeiten in 17 Parametern der mittleren Struktur des Langzeitspektrums im Frequenzbereich von 1000 bis 5000 Hz zugrunde; ansonsten gelten dieselben Erläuterungen wie bei Abbildung 16.

Betrachtet man die Imitationen, so fällt folgendes auf (Abbildung 16 und 17).

- (1) Fast alle Imitationen liegen außerhalb des normalen Streuungsbereichs von (A) (obwohl sie zum größten Teil unter denselben technischen Bedingungen aufgenommen wurden wie die Originalstimme des Imitators).
- (2) Bei keiner der Imitationen gelingt es dem Imitator, in den Streuungsbereich der angestrebten Originalstimme einzudringen (mit Ausnahme von (E) in Abbildung 17, wenn man die eine weit von den anderen beiden entfernte Probe in den normalen Streuungsbereich einbezieht; vermutlich handelt es sich hier aber um einen Geräteartefakt).
- (3) Bei Verwendung des gesamten Langzeitspektrums für die Ähnlichkeitsvergleiche (Abbildung 16) zeigt sich eine schwache Tendenz, daß Sprecher (A) sich beim Imitieren auf die angestrebten Originale zubewegt. Bezüglich der mittleren Struktur (Abbildung 17) gelingt ihm dies nur bei zwei Sprechern; von (C) bewegt er sich sogar weg.
- (4) Die intraindividuelle Variabilität des Imitators beim Imitieren ist noch deutlich geringer als die interindividuelle Variabilität der Originalstimmen, obwohl es ja die Intention des Imitators war, letztere zu erreichen.
- (5) Imitationen verschiedener angestrebter Originale liegen zum Teil sehr nah beieinander, obwohl sie für das menschliche Ohr so unterschiedlich klingen wie die Stimmen verschiedener Sprecher.

Als ein weiterer Test wurde mit Hilfe der Diskriminanzanalyse (Prozedur "discriminant" aus SPSS, Beutel und Schubö 1983) geprüft, inwieweit der Imitator beim Imitieren noch als er selbst identifizierbar ist.

Für die Aufstellung der Diskriminanzfunktionen wurden von den 17 Werten des mittleren Struktur des Langzeitspektrums 7 ausgewählt, bei denen die interindividuelle Varianz erheblich größer ist als die intraindividuelle ($P < 0.001$). Eine größere Zahl von Parametern kann aufgrund der begrenzten Anzahl von Freiheitsgraden bei der gegebenen Stichprobengröße nicht verwendet werden.

Da bei einer systematischen Vorauswahl von Variablen die Interpretation von Signifikanztests bei der Diskriminanzanalyse Schwierigkeiten bereitet, wurde rein empirisch vorgegangen: In der Analysephase wurden sämtliche Originalstimmen von (A), (B), (C), (D) und (E) eingegeben; als Gruppe wurden jeweils die Proben eines Sprechers bezeichnet. Die Imitate wurden dann in der Zuordnungsphase der Gruppe zugeordnet, in der die Zugehörigkeitswahrscheinlichkeit am höchsten lag.

Von den 12 Imitationen konnten 4 dem Sprecher (A) zugeordnet werden, die Identifikationsrate liegt also bei 33% und somit noch leicht über dem Zufall (20% bei 5 Sprechern). Nur in zwei Fällen allerdings wurde der Imitator für den Imitierten gehalten;

das entspricht 25% der Falschidentifikationen und so genau der Zufallserwartung (für Falschidentifikationen kommen nur 4 Möglichkeiten in Frage).

Gibt man nun in der Analysephase nicht nur Originalstimmen sondern auch einige Imitationen ein und gibt so dem System die vergrößerte intraindividuelle Variabilität des Imitators zu erkennen, müßte die Erkennungsrate ansteigen.

Um dies zu testen, wurden zusätzlich zu den Originalen (B), (C), (D) und (E) 19 Proben von (A) verwendet (Imitationen und Originalstimmen zusammen). Es wurden zwei Diskriminanzanalysen gerechnet, einmal wurden 10 Proben von (A) in der Analysephase weggelassen und nur für die Zuordnung verwendet das andere Mal die restlichen 9.

Von den insgesamt 19 Zuordnungen waren 15 richtig; das entspricht einer Erkennungsrate von fast 80%.

Aus dem Imitationsexperiment lassen sich nun folgende Ergebnisse zusammenfassen:

Ein geübter Sprecher kann seine normale Variabilität im Langzeitspektrum willentlich erheblich erhöhen. Dies ist jedoch nicht ganz so einfach für ihn; der Imitator sagte, er müsse dafür auch die Körperhaltung und die Gestik des Imitierten nachahmen, sich also ganzheitlich in ihn hineinversetzen.

Bezüglich des auditiven Eindrucks gelingt es ihm sehr gut, die Stimme des Imitierten zu treffen; im Langzeitspektrum nicht besonders gut. Bezüglich der mittelfeinen Struktur des Langzeitspektrums kann keinerlei Tendenz festgestellt werden, daß der Imitator dem Imitierten ähnlicher wird.

Als Erklärung dafür kann man anführen, daß es dem Imitator ja genügt, wenn er im menschlichen Zuhörer den Eindruck eines anderen Sprechers erweckt. Für eine Annäherung an die angestrebte Stimme in objektiven Parametern fehlt ihm außerdem beim Training ein geeignetes "feed back", an dem er sich orientieren könnte. Ob er mit Hilfe eines solchen feed backs allerdings in die objektiv meßbare Variabilität der Originalstimmen gelangen kann oder ob es anatomische oder physiologische Grenzen gibt, die dies verhindern, kann hier nicht entschieden werden.

Die Diskrepanz zwischen auditivem Eindruck und objektiven Stimmkriterien zeigt, daß der menschliche Zuhörer normalerweise andere Kriterien benutzt als den "Durchschnittsklang" (als den man ein auditives Korrelat des Langzeitspektrums bezeichnen könnte) für die Erkennung eines Sprechers. Diese Kriterien liegen vermutlich im Zeitverlauf des Sprachsignals und in Frequenzmustern sehr kurzer Signalstücke (z. B. Lautübergänge) und können daher nur auf sprachsegmentaler Ebene analysiert werden.

Aus der Beobachtung, daß sich auch in Merkmalen, die offensichtlich für den auditiven Eindruck und daher auch für die Intention des Imitators wenig relevant sind, beim Imitieren die intraindividuelle Variabilität sehr stark erhöht, ergibt sich eine praktische Konsequenz für die automatische Sprecheridentifizierung: Ein Sprecher, von dem man nicht genau weiß, ob er seine Stimme absichtlich verstellt hat, und von dem man nur Referenzproben mit seiner normalen Stimme hat, ist nicht identifizierbar.

3.8. Methode zur Erfassung des Langzeitspektrums im Frequenzbereich unterhalb 1000 Hz mit höherer Frequenzauflösung

3.8.1. Methode

Bei der Betrachtung der Langzeitspektren der bisherigen Untersuchungen war aufgefallen, daß der Frequenzbereich unterhalb 1000 Hz eine sehr große Variabilität aufweist. Für eine sinnvolle Interpretation dieser Variabilität ist jedoch aus folgendem Grund eine bessere Frequenzauflösung erforderlich: Im Bereich zwischen ca. 80 und 300 Hz liegt bei Erwachsenen die Grundfrequenz der Sprechstimme. Es ist daher zu erwarten, daß die Energieverteilung im Spektralbereich unterhalb 1000 Hz im wesentlichen durch die mittlere Grundfrequenz und ihre Harmonischen geprägt wird. Da aber interindividuelle Differenzen in der Grundfrequenz oft weit weniger als 40 Hz betragen, ist eine Frequenzauflösung von ca. 10 Hz erforderlich. Eine noch feinere Frequenzauflösung dürfte insofern keine wesentliche Verbesserung der Darstellung bringen, als der Grundfrequenzgipfel aufgrund von Schwankungen in der Grundfrequenz beim Sprechen (Intonation, "intrinsic pitch") im Langzeitspektrum nicht so scharf erscheint wie in einem Linienspektrum, dem ein streng periodisches Signal zugrundeliegt; bei noch höherer Frequenzauflösung wäre dann eine hohe Korrelation zwischen Nachbarwerten zu beobachten und somit eine Redundanz in den Ergebnisdaten, die nur Speicherplatz verbraucht. Außerdem würden dann die den Einzelspektren zugrundeliegenden Signalstücke zu lang werden und man bräuchte zur Erstellung einer stabilen Langzeitstruktur sehr lange Textpassagen.

Die gesonderte Betrachtung des Frequenzbereichs unterhalb 1000 Hz bietet sich insofern an, als bei allen untersuchten Sprechern im Bereich zwischen ca. 900 und 1000 Hz ein **ausgeprägtes Energieminimum zu beobachten ist.**

Die Berechnung des Grundfrequenzspektrums (das Langzeitspektrum im Bereich

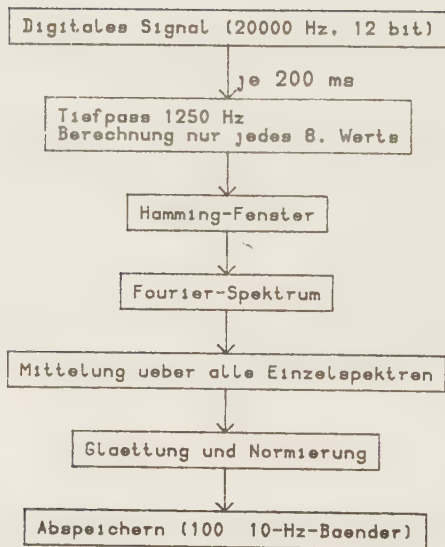


Abbildung 18:

Berechnung des Grundfrequenzspektrums

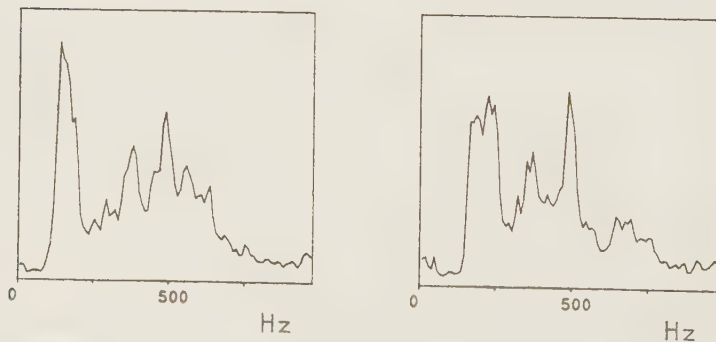


Abbildung 19:

Zwei Grundfrequenzspektrn (links männlich, rechts weiblich);

der Unterschied in der mittleren Grundfrequenz ist deutlich an der Lage des ersten hohen Gipfels zu erkennen.

(Die mittlere Grundfrequenz beträgt bei Männern im Schnitt ca. 120 Hz, bei Frauen ca. 200 Hz. Sie wird von den Stimmlippen im Kehlkopf erzeugt und entspricht ungefähr dem, was wir als Sprechtonhöhe wahrnehmen.)

unterhalb 1000 Hz sei hier als Grundfrequenzspektrum bezeichnet) wurde in folgenden Schritten vorgenommen (siehe Abbildung 18):

- (1) Das Signal, das mit einer Samplingrate von 20000/s digital vorliegt, wird tiefpaßgefiltert (Schnittfrequenz 1250 Hz, entsprechend 1/16 der Sampling-Frequenz); vom gefilterten Signal wird nur jeder 8. Wert verwendet. (Diese Datenreduktion ist erforderlich, da sonst bei der folgenden Fourier-Transformation sehr lange Rechenzeiten entstehen würden.)
- (2) Das reduzierte Signal wird nun in Stücken von 512 Werten (das entspricht ca. 0.2 Sekunden Sprache) nach Bewertung mit einem Hamming-Fenster Fourier-transformiert, so daß (nach Eliminierung der Phasenbeziehungen) Amplitudenspektren mit 256 Werten von 5 bis 1250 Hz entstehen.
- (3) Die Einzelspektren werden aufsummiert, das Summenspektrum wird normiert.
- (4) Nachdem sich herausgestellt hat, daß sich auch die Grundfrequenzspektren durch Glättung analog der in Abbildung 7 (Seite 37) dargestellten nicht wesentlich verändern, wurden sie ebenfalls geglättet; es wurden 100 Werte abgespeichert, entsprechend 100 10-Hz-Bändern von 10 bis 1000 Hz (oberhalb 1000 Hz genügt das 80 Hz-Auflösungsvermögen des "normalen" Langzeitspektrums).

Abbildung 19 zeigt zwei Grundfrequenzspektren; die Lage der mittleren Grundfrequenz ist deutlich am ersten ausgeprägten Gipfel zu erkennen.

3.8.2. Sprechlautstärke und Tonhöhe

Die in 3.5. dargestellten Einflüsse von Lautstärke und Tonhöhe auf das Langzeitspektrum beinhalten im wesentlichen nur eine Anhebung der Intensität im Bereich von 2000 bis 3000 Hz bei laut gesprochenen Proben. Insbesondere Veränderungen im Grundfrequenzbereich konnten aufgrund der groben Frequenzauflösung nicht gut herausgearbeitet werden.

In Abbildung 20 sind die Grundfrequenzspektren der vier Sprecher dargestellt; von jedem Sprecher sind alle fünf Proben jeweils in einem Diagramm abgebildet (von unten nach oben: leise, normal, laut, tief und hoch gesprochen).

Im Bereich unterhalb 300 Hz ist bei allen Spektren der Grundfrequenzgipfel (mehr oder weniger ausgeprägt) erkennbar. Ein Vergleich der hoch und der tief gesprochenen Proben (jeweils die oberen beiden) zeigt, daß der Grundfrequenzgipfel bei hoch gesprochenen Proben deutlich nach rechts verlagert ist.

Daß die Grundfrequenz auch mit der Sprechlautstärke angehoben wird, zeigt sich in den

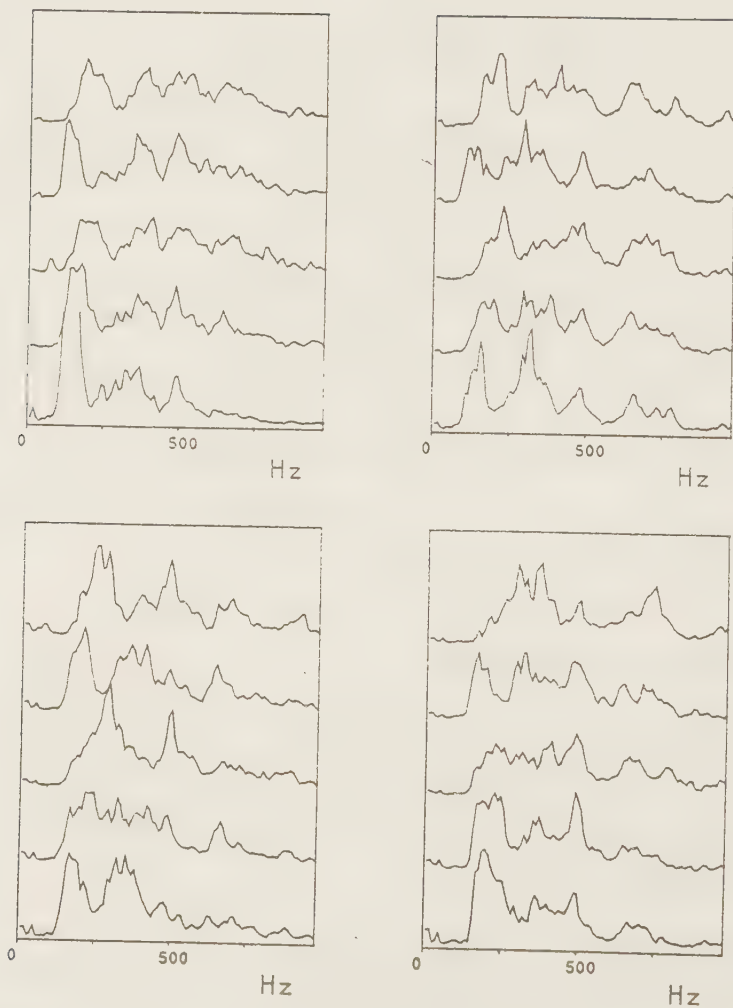


Abbildung 20:

Jedes der vier Diagramme zeigt fünf Grundfrequenzspektren eines Sprechers; die fünf zugrundeliegenden Sprechproben wurden mit (von unten nach oben) leiser, normaler, lauter, tiefer und hoher Stimme gesprochen.



Abbildung 21:

Zweidimensionale Darstellung der Ähnlichkeitsbeziehungen im Frequenzbereich von 300 bis 1000 Hz im Grundfrequenzspektrum:

Buchstaben: Sprecher

Zahlen: 1 = leise, 2 = normal, 3 = laut, 4 = tief, 5 = hoch.

Es ist weder eine Gruppierung nach Sprechern noch eine nach Sprechweise zu beobachten.

drei unteren Linienzügen jedes Diagramms.

Auffallend ist, daß die Kurven oberhalb 300 Hz nicht nur aus "Harmonischen" der Grundfrequenz bestehen sondern sehr unregelmäßig verlaufen; offensichtlich wird also der Verlauf in diesem Frequenzbereich eher von Eigenschaften der Resonanzräume als von der Grundfrequenz selbst geprägt.

Um zu sehen, ob sich in diesem Bereich stabile Sprechermerkmale ausdrücken, wurde wieder ein Ähnlichkeitsvergleich durchgeführt. Jede Probe wurde in den 70 Werten oberhalb 300 Hz mit jeder verglichen; die Ähnlichkeitsbeziehungen wurden in zwei Dimensionen dargestellt (siehe 3.5.2., Seite 45).

Wie Abbildung 21 zeigt, findet keine Gruppierung nach Sprechern statt; es bietet sich auch keine einfache Interpretation der beobachtbaren Anordnung an.

Offensichtlich verändert sich der Spektralverlauf in diesem Frequenzbereich empfindlich mit Sprechtonhöhe und Lautstärke, ohne daß diese beiden Einflußgrößen hier unmittelbar abgebildet werden. Auch in Abbildung 20 kann man keine Regelmäßigkeiten erkennen, nach denen Lautstärke und Tonhöhe im Frequenzbereich oberhalb 300 Hz zum Ausdruck kommen.

3.8.3. Geschlechtsdimorphismus

Nachdem, wie allgemein bekannt ist, Frauen im Schnitt höhere Stimmen haben als Männer, müßte dieser Geschlechterunterschied im Grundfrequenzspektrum zum Ausdruck kommen.

Um dies zu zeigen, wurde eine Stichprobe von 10 männlichen und 10 weiblichen Versuchspersonen untersucht; es wurden von ca. 20 Sekunden aus "Der Nordwind und die Sonne" Grundfrequenzspektren erstellt und die Ähnlichkeitsbeziehungen zweidimensional aufgezeichnet; das Geschlecht wurde nachträglich markiert.

Wie man in Abbildung 22 (links oben) sieht, ist der Geschlechtsdimorphismus im Grundfrequenzspektrum sehr stark ausgeprägt. Läßt man nur Teile des Grundfrequenzspektrums in die Ähnlichkeitsvergleiche eingehen, so kann man sehen, in welchen Frequenzbereichen der Geschlechtsdimorphismus am stärksten ausgeprägt ist. Der Bereich unterhalb 300 Hz (rechts oben), in dem ja die Grundfrequenz selbst liegt, trennt die Geschlechter noch vollständig. Im Bereich von 300 bis 1000 Hz (links unten) gibt es bereits Überlappungen; im Bereich von 500 bis 1000 Hz (rechts unten) ist kein Geschlechtsdimorphismus zu beobachten.

Letzteres spricht auch wieder dafür, daß der Spektralverlauf in diesem Bereich nicht unmittelbar von der Grundfrequenz geprägt wird.



Abbildung 22:

Geschlechtsdimorphismus im Grundfrequenzspektrum:

Die Ähnlichkeiten (Korrelationskoeffizient als Ähnlichkeitsmaß) in verschiedenen Frequenzbereichen des Grundfrequenzspektrums wurden mit Hilfe der smallest space analysis zweidimensional dargestellt.

links oben: gesamtes Grundfrequenzspektrum (10 - 1000 Hz)

rechts oben: Frequenzbereich 10 - 300 Hz

links unten: Frequenzbereich 300 - 1000 Hz

rechts unten: Frequenzbereich 500 - 1000 Hz

Geschlechtermarkierung: Pfeil = männlich, Mond = weiblich.

3.9. Erfassung sprecherbezogener Parameter aus dem Sprechrhythmus

Beschränkt man sich auf die Untersuchung von Langzeitspektren, so hat man sämtliche Informationen, die in Form von Zeitverlaufsmodellen im Sprachsignal stecken (z. B. Setzen der Sprechpausen, Betonungen, Grundfrequenzverlauf ...), nicht erfaßt.

Nachdem diese Zeitmuster aber in viel stärkerem Maße vom gesprochenen Text abhängen als der "Durchschnittsklang" der Sprechstimme, dürfte es einige Schwierigkeiten bereiten, hier sprecherbezogene Parameter zu extrahieren. Außerdem ist der Sprechrhythmus wesentlich leichter vom Sprecher willentlich zu verändern als der Klang, so daß bei der Erfassung stabiler, sprechertypischer Eigenschaften eine große situationsbedingte Unschärfe zu erwarten ist.

Zwei wesentliche Derivate des Sprachsignals, in denen der Sprechrhythmus zum Ausdruck kommt, sind der Intensitäts- und der Grundfrequenzverlauf (siehe Abbildung 23); auf sie sei die folgende Untersuchung beschränkt.

Es soll hier versucht werden, in diesen beiden Zeitsignalen Periodizitätsmuster zu finden, in denen, zumindest grob statistisch, Sprechereigenschaften zum Ausdruck kommen. Hierfür werden Langzeitspektren im Bereich der Rhythmuswahrnehmung (ca. 0,3 bis 10 Hz) erstellt. Da diese Frequenzmuster in hohem Maße von der Sprechgeschwindigkeit abhängen, wird versucht, durch Dehnung oder Kompression des Signals zumindest den linearen Einfluß der Sprechgeschwindigkeit zu eliminieren.

3.9.1. Intensitätsverlauf

Für die Aufzeichnung des Intensitätsverlaufs wird das Signal quadriert und tiefpaßgefiltert. Da sich ein tiefpaßgefiltertes Signal mit einer geringeren Samplingrate darstellen läßt als das Ausgangssignal, muß nicht jeder Wert berechnet werden.

Die Filterung wurde realisiert, indem jeweils n Werte aufaddiert wurden; es wurde jeder n -te Wert berechnet.

Durch lineare Zeitkompression soll nun erreicht werden, daß von jeder Sprechprobe dieselbe Textstelle mit gleich vielen Intensitätswerten dargestellt wird. Hierfür wurde n einfach von der Sprechgeschwindigkeit abhängig gemacht.

Anfang und Ende der Sprechprobe wurden durch empirisch ermittelte Schwellenwerte bestimmt. Von einer Textstelle, die im Schnitt etwa 20 Sekunden Redezeit benötigt, wurden immer 4096 Intensitätswerte berechnet. Da sich bei Sprechern mit tiefen Stimmen bei dieser Zeitaufösung die Grundfrequenz noch auf den Verlauf auswirkt, wurde die

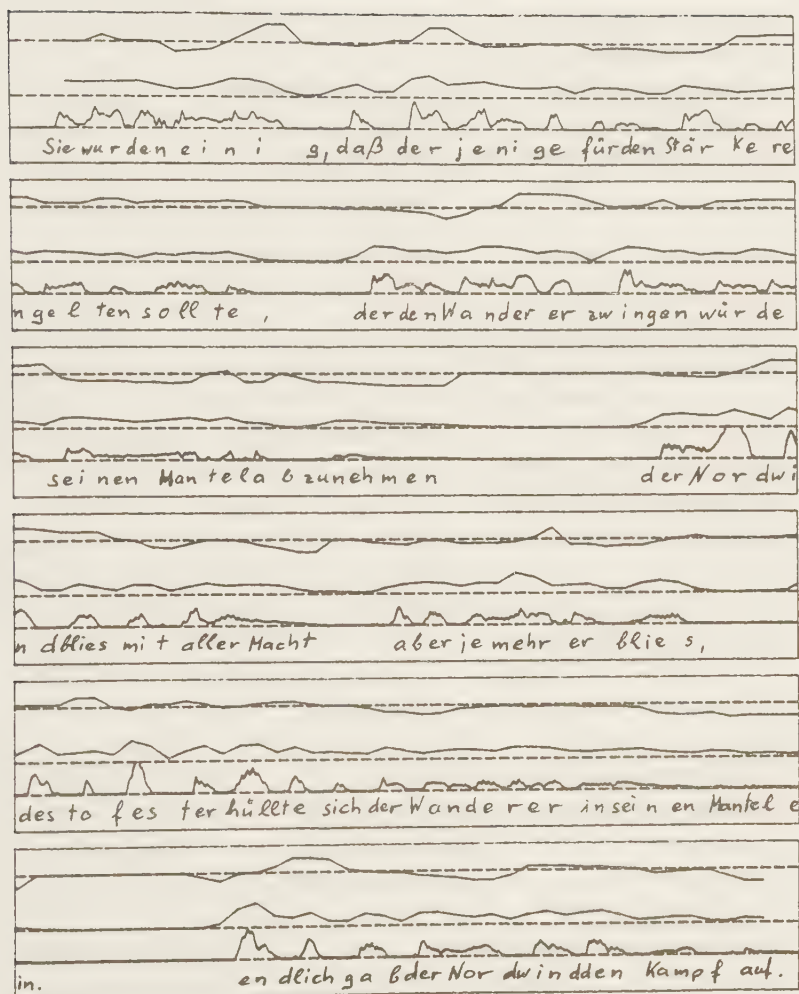


Abbildung 23:

Grundfrequenz- und Intensitätsverlauf einer Sprechprobe von ca. 20 Sekunden:

obere Kurve: Grundfrequenzverlauf (256 Werte)

mittlere Kurve: Verlauf der Intensität der dem Grundfrequenzverlauf zugrundeliegenden Einzelspektren (256 Werte)

untere Kurve: Intensitätsverlauf nach Quadrieren und Tiefpaßfiltern der Sprechprobe (4096 Werte)

Verlaufskurve zusätzlich mit einem 18-stufigen nichtrekursiven Tiefpaß, der bei einem Viertel der Samplingfrequenz steil schneidet, gefiltert. In Abbildung 23 (untere Kurve) ist ein solcher Intensitätsverlauf dargestellt.

Zur Ermittlung eines nicht zeitkomprimierten Intensitätsverlaufs wurde $n=117$ fest eingestellt. (Diese Zahl wurde aus praktischen Gründen so gewählt, damit mit 4096 Werten 24 Sekunden Sprache dargestellt werden können.)

3.9.2. Grundfrequenzverlauf

Die Aufzeichnung des Grundfrequenzverlaufs bereitet wesentlich mehr Probleme als die des Intensitätsverlaufs. Zur digitalen Extraktion der Grundfrequenz aus dem Sprachsignal gibt es zwar eine Vielzahl von Algorithmen; keiner arbeitet aber bei allen Sprechern fehlerfrei. Außerdem ist der Umgang mit Sprechpausen und stimmlosen Passagen schwierig, da hier eine automatische Grundfrequenzbestimmung sinnlose Werte liefert, die als solche erkannt werden müssen.

Im Rahmen dieser Arbeit wurde ein speziell auf die Fragestellung zugeschnittenes Verfahren entwickelt, das den Grundfrequenzverlauf einer Sprechprobe von ca. 20 Sekunden Dauer mit 256 Werten so darstellt, daß man daran den Text verfolgen kann (siehe Abbildung 23, obere Kurve).

Die Berechnung wurde folgendermaßen durchgeführt:

- (1) Digitale Bandpaßfilterung, die nur Frequenzen zwischen ca. 60 und 600 Hz durchläßt und Datenreduktion (neue Samplingrate: 1250 Hz);
- (2) Bestimmung von Anfang und Ende der Sprechprobe;
- (3) Aufteilung der Probe auf 256 überlappende Stücke mit je 256 Werten (der Grad der Überlappung ist wieder von der Sprechgeschwindigkeit abhängig, zur Erlangung einer linearen Zeitkompression)
- (4) Berechnung eines Einzelspektrums von jedem der 256 Signalstücke mit Hilfe der schnellen Fourier-Transformation (ein Spektrum besteht aus 128 5-Hz-Bändern von 5 bis 640 Hz). Die ersten 40 Einzelspektren einer Sprechprobe sind in Abbildung 24 dargestellt.
- (5) Ermittlung der Grundfrequenz aus jedem Einzelspektrum:
 - Ist die Gesamtintensität im Spektrum zu gering, so wird ein fehlender Wert registriert (Sprechpauseneliminierung).
 - Ermittlung des ersten Gipfels von links, der einen (empirisch ermittelten) Schwellenwert übersteigt.

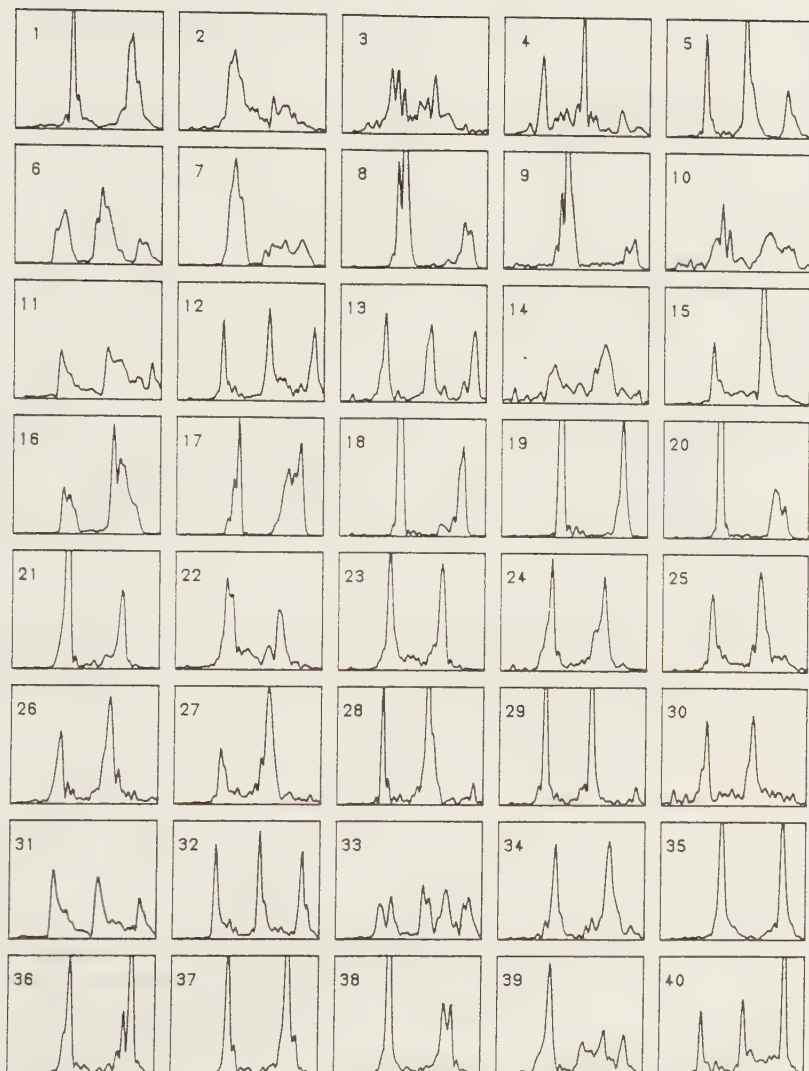


Abbildung 24:

40 Einzelspektren, aus denen die ersten 40 Werte der Grundfrequenzverlaufskurve ermittelt werden (Abszisse: Frequenz, 0 bis 600 Hz; Ordinate: Intensität)

- Ist bei der doppelten Frequenz dieses Gipfels (unterhalb 150 Hz auch bei der dreifachen) ebenfalls eine erhöhte Intensität vorhanden, so wird die Frequenz bei diesem Gipfel als Grundfrequenz akzeptiert. Werden die Bedingungen nicht erfüllt, so wird es mit dem nächsten Gipfel von links probiert.
 - Läßt sich kein Gipfel finden, der den Bedingungen genügt, so wird ein fehlender Wert registriert.
- (6) Korrektur des Grundfrequenzverlaufs:
- Ermittlung des am häufigsten auftretenden Werts.
 - Werte, die über dem 1.75-fachen des häufigsten Werts liegen, werden halbiert (es wurde mit hoher Wahrscheinlichkeit die erste Oberschwingung für die Grundfrequenz gehalten).
 - Werte, die sich zu stark von ihren Nachbarwerten unterscheiden, werden als fehlend betrachtet.
 - Alle fehlenden Werte werden, falls beide Nachbarwerte vorhanden sind, durch ihr arithmetisches Mittel, ansonsten durch den am häufigsten auftretenden Wert ersetzt.
 - Die Gesamtzahl der fehlenden Werte wird registriert, als Maß für die Plausibilität der Grundfrequenzkurve.

Dieses Verfahren wurde an einer Stichprobe von 20 Sprechern getestet, bei allen lieferte es plausible Verlaufskurven (wie in Abbildung 23); die Gesamtzahl der fehlenden Werte lag ungefähr bei einem Drittel aller Werte; fast alle fehlenden Werte konnten auditiv wahrnehmbaren Sprechpausen zugeordnet werden.

3.9.3. Berechnung von Rhythmusspektren aus Intensitäts- und Grundfrequenzverlauf

Zur Berechnung von Langzeit- Rhythmusspektren wurden die Verlaufskurven auf 29 zu 75% überlappende Stücke von $1/8$ Redezeit aufgeteilt. Die Stücke wurden mit einem Hamming-Fenster bewertet und Fourier-transformiert; die so gewonnenen Einzel-Rhythmusspektren wurden aufsummiert, das Summenspektrum dann auf gleiche Gesamtintensität normiert.

Es wurden folgende drei Rhythmusspektren berechnet:

- (1) RSPKI: Rhythmusspektrum des Intensitätsverlaufs mit absoluter Frequenzskala (ohne lineare Zeitkompression), 20 $1/3$ Hz-Bänder, von $1/3$ bis 6 Hz.
- (2) RSPKIZ: Rhythmusspektrum des Intensitätsverlaufs nach linearer Zeitkompression der Verlaufskurve, 10 Frequenzbänder von $8/\text{Lesezeit}$ bis $80/\text{Lesezeit}$. Diese Frequenzauflösung entspricht bei einer Lesezeit von 20 Sekunden $1/4$ Hz.

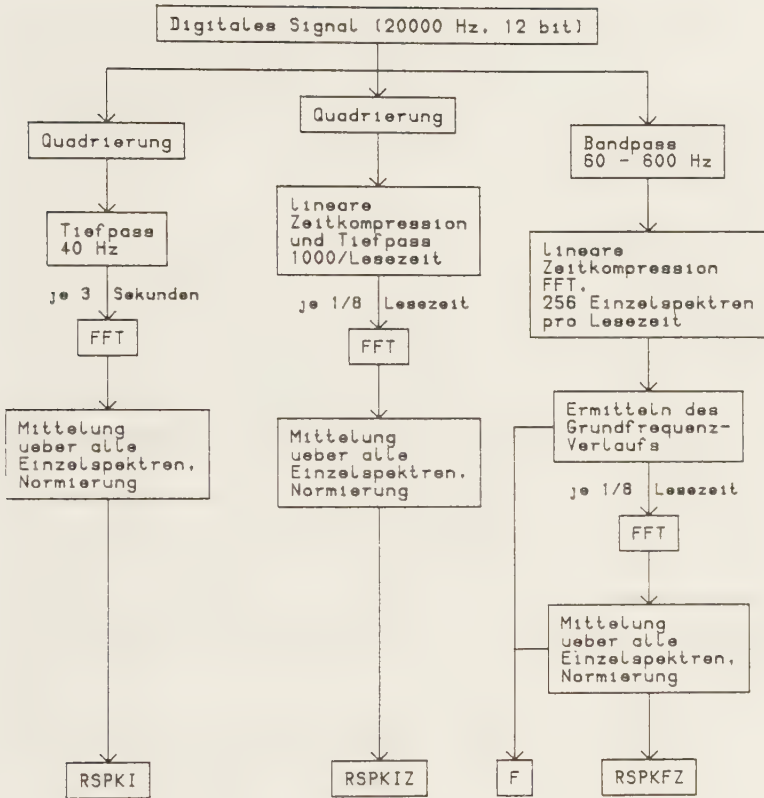


Abbildung 25:

Berechnung der Rhythmusspektren und zusätzlicher Zahlenwerte:

FFT: Bewertung mit Hamming - Fenster, schnelle Fourier - Transformation und Phaseneeliminierung

RSPKI: Rhythmusspektrum des Intensitätsverlaufs (20 1/3 Hz - Bänder)

RSPKIZ: Rhythmusspektrum des Intensitätsverlaufs nach linearer Zeitkompression (10 Frequenzbänder von 8/Lesezeit)

RSPKFZ: Rhythmusspektrum des Grundfrequenzverlaufs nach linearer Zeitkompression (10 Frequenzbänder von 8/Lesezeit)

F: 6 Zahlenwerte mit folgender Bedeutung: F1 = Lesezeit, F2 = mittlere Grundfrequenz, F3 = relative Standardabweichung der Grundfrequenz, F4 = Veränderungsdynamik benachbarter Grundfrequenzwerte, F5 = Dynamik langweiliger Grundfrequenzschwankungen (über 2 s), F6 = Dynamik kurzweiliger Grundfrequenzschwankungen (1/4 bis 2 s)

(3) RSPKFZ: Rhythmusspektrum des Grundfrequenzverlaufs nach linearer Zeitkompression der Verlaufskurve, ebenfalls 10 Frequenzbänder von 8/Lesezeit bis 80/Lesezeit.

Da bei Mittelung über nur so wenige Einzelspektren noch keine sehr stabile, textunabhängige Langzeitstruktur erwartet werden kann, wurde bei allen Sprechern dieselbe Textstelle untersucht.

Zusätzlich zu diesen drei Rhythmusspektren wurden noch sechs Einzelwerte berechnet:

F1: Lesezeit;

F2: mittlere Grundfrequenz aus der Grundfrequenzverlaufskurve;

F3: Standardabweichung der Grundfrequenz, relativ zu F2;

F4: Wurzel aus der mittleren quadrierten Differenz benachbarter Grundfrequenzwerte, dividiert durch F2 (als Maß für die Veränderungsdynamik der Grundfrequenz in kurzen Zeiträumen);

F5: Standardabweichung der häufigsten Grundfrequenz in den den Rhythmus-Einzelspektren zugrundeliegenden Signalstücken (hier soll die Dynamik sehr langwelliger (mehr als 2 Sekunden) Grundfrequenzschwankungen erfaßt werden);

F6: Gesamtintensität des Langzeit-Rhythmusspektrums vor der Normierung, als Maß für die Veränderungsdynamik der Grundfrequenz innerhalb von Zeiträumen bis ca. 2 Sekunden;

In Abbildung 25 ist die Berechnung der Rhythmusspektren schematisch dargestellt.

4. Studie an einer größeren Stichprobe von Sprechern zur Erfassung akustischer Eigenschaften und anthropologischen, psychologischen und auditiven Korrelaten

4.1. Datenmaterial

Diese Untersuchung wurde an 117 Versuchspersonen durchgeführt; fast alle sind Biologiestudenten und –studentinnen (70 weiblich, 47 männlich) im Alter zwischen 20 und 30 Jahren.

Das von jeder Person erfaßte Material setzt sich folgendermaßen zusammen:

- (1) eine Tonbandaufnahme von "Der Nordwind und die Sonne",
- (2) ein Stimm – Selbstbeurteilungsfragebogen (siehe Anhang, Seite A3),
- (3) ein kurzer Persönlichkeitsfragebogen (Eysenck personality inventory, EPI) mit 57 ja/nein – Antworten,
- (4) 27 anthropometrische Merkmale (siehe Anhang, Seite A1),
- (5) Lebensalter und Geschlecht

Die auf Tonband vorliegenden Sprechproben wurden sowohl akustisch als auch auditiv ausgewertet.

4.1.1. Akustische Analyse der Sprechproben

Die Tonaufnahmen wurden mit einem Geraet Uher Report Monitor 4200 und dem dazugehörigen Mikrophon Uher 518a, Mund – Mikrophonabstand 30 cm, Bandlaufgeschwindigkeit 19 cm/s, erstellt. Es wurde auf Konstanz der Aufnahmebedingungen geachtet (Raumakustik, Aufnahmewinkel zur Sprechrichtung).

Die Sprechproben wurden digitalisiert und auf Magnetband abgespeichert (siehe Abbildung 6, Seite 37). Hierfür wurde von jedem Sprecher die in Abbildung 23 (Seite 63) aufgezeichnete Textstelle verwendet.

Für die Berechnung akustischer Parameter wurden die in den Voruntersuchungen beschriebenen Verfahren zur Berechnung von Langzeitspektrum (Abbildung 1, Seite 27, mit den auf Seite 36 beschriebenen Modifikationen), Grundfrequenzspektrum (Abbildung 18, Seite 56) und Rhythmuspektren (Abbildung 25, Seite 67) zu einem Programmpaket zusammengefaßt.

Folgende akustischen Parameter wurden von jeder Sprechprobe berechnet und

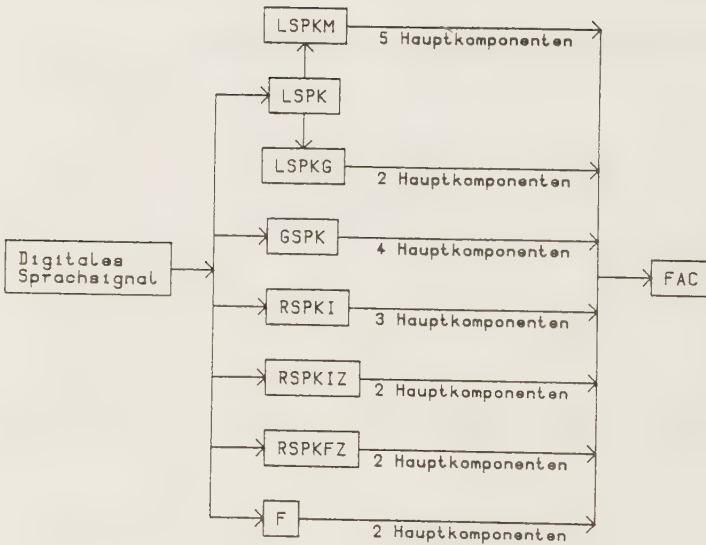


Abbildung 26:

Extraktion akustischer Parameter aus dem Sprachsignal

LSPK: Langzeitspektrum, 80 – 8000 Hz

LSPKM: mittlere Struktur des Langzeitspektrums, 1000 – 6000 Hz

LSPKG: Grobstruktur des Langzeitspektrums, 80 – 8000 Hz

GSPK: Grundfrequenzspektrum, 10 – 1000 Hz

RSPKI: Rhythmusspektrum des Intensitätsverlaufs, 1/3 – 6 Hz

RSPKIZ: Rhythmusspektrum des Intensitätsverlaufs nach linearer Zeitkompression

RSPKFZ: Rhythmusspektrum des Grundfrequenzverlaufs nach linearer Zeitkompression

F: Lesezeit (F1), Grundfrequenz (F2) und 4 Streuungsmaße der Grundfrequenz (F3 – F6)

FAC: 20 durch Hauptkomponentenanalysen gewonnene Faktoren

abgespeichert:

LSPK: Langzeitspektrum, 100 80-Hz-Bänder im Frequenzbereich von 80 bis 8000 Hz,

LSPKG: Grobverlauf des Langzeitspektrums, 8 1000-Hz-Bänder,

LSPKM: Mittlere Struktur des Langzeitspektrums, 22 Werte im Bereich von 1000 bis 6000 Hz, Frequenzauflösung 240 Hz,

GSPK: Grundfrequenzspektrum, 100 10-Hz-Bänder im Frequenzbereich von 10 bis 1000 Hz,

RSPKI: Rhythmusspektrum des Intensitätsverlaufs, 20 $1/3$ -Hz-Bänder im Frequenzbereich von $1/3$ bis 6 Hz,

RSPKIZ: Rhythmusspektrum des Intensitätsverlaufs nach linearer Zeitkompression der Sprechprobe, 10 Frequenzbänder von 8/Lesezeit bis 80/Lesezeit,

RSPKfZ: Rhythmusspektrum des Grundfrequenzverlaufs nach linearer Zeitkompression der Sprechprobe, 10 Frequenzbänder von 8/Lesezeit bis 80/Lesezeit,

F: 6 Einzelwerte: Lesezeit, mittlere Grundfrequenz und vier Streuungsmaße für den Grundfrequenzverlauf (siehe Seite 67).

Die Gesamtzahl der Parameter beträgt also 276.

Da einige statistische Verfahren aber nur sinnvoll interpretierbare Ergebnisse liefern, wenn die Zahl der eingehenden Parameter wesentlich kleiner ist als die Größe der Stichprobe, ist es angebracht, die Datenmenge noch zu reduzieren.

Aus theoretischen Überlegungen heraus ist dies auch sinnvoll, da in den Daten, sofern man an der interindividuellen Variabilität interessiert ist, eine hohe Redundanz zu erwarten ist. Man kann sich diese Redundanz als eine Information vorstellen, die bei allen Sprechproben dieselbe ist, also zum Beispiel die Information, daß es sich um ein menschliches Sprachsignal handelt und nicht um Motorenlärm, Vogelgezwitscher oder das Rauschen eines Gebirgsbachs.

Möchte man die interindividuelle Varianz, die in den Daten steckt, mit möglichst wenigen Parametern beschreiben, so bietet sich die Hauptkomponentenanalyse als Verfahren zur Datenreduktion an. Man kann sie sich geometrisch veranschaulichen: sie ist eine Koordinatentransformation, die den n -dimensionalen nichtorthogonalen (es gibt viele Interkorrelationen unter den n Merkmalen) Merkmalsraum zunächst in einen orthogonalen Raum mit n Dimensionen so überführt, daß die erste Dimension die ist, in der die Streuung am größten ist, die zweite die, die orthogonal zur ersten die höchste Streuung aufweist ...

Ist man bereit, auf einen relativ kleinen Varianzanteil zu verzichten, so kann man, je nach Redundanz in den Ausgangsdaten, sehr viele Dimensionen vernachlässigen und damit die Zahl der zu speichernden Werte reduzieren.

Aus den akustischen Parametern wurden so viele Faktoren berechnet, daß ungefähr 80% der Varianz damit beschrieben wird; die verschiedenen Teil-Datensätze wurden getrennt faktorisiert. Insgesamt wurden 20 Faktoren berechnet; ihre Zusammensetzung geht aus Abbildung 26 hervor. Zur Darstellung des Geschlechtsdimorphismus wurden die Faktoren aus allen Proben gemeinsam berechnet, zur Darstellung der übrigen Zusammenhänge wurden sie bei beiden Geschlechtern getrennt extrahiert.

Der Vorteil der Datenreduktion ist zum Teil damit erkauft, daß es sich bei den Faktoren um unanschauliche Zahlenwerte handelt. Während man sich noch einigermaßen vorstellen kann, wie eine Stimme klingt, bei der ein gewisser Frequenzbereich eine hohe Intensität aufweist, so kann man sich, selbst wenn man das Ladungsmuster der Faktoren dabei betrachtet, nur schwer vorstellen, wie eine Stimme klingt, die im Faktor 3 einen hohen Wert aufweist. Es ist also nicht sinnvoll, auf die Ausgangsdaten zugunsten der Faktoren vollständig zu verzichten.

4.1.2. Anthropometrische Daten

Von jeder Versuchsperson wurden Alter und Geschlecht registriert und 27 anthropometrische Meßwerte (K1 bis K27) aufgenommen. Die Auswahl der Maße richtete sich im wesentlichen nach Knußmann (1968); vor allem wurden die ausgewählt, in denen der Geschlechtsdimorphismus zum Ausdruck kommt und die, die zur Berechnung der im Kapitel "Konstitution" angegebenen Diskriminanzfunktion (bezüglich Pykno/Leptomorphie) erforderlich sind.

Eine Beschreibung der Merkmale findet sich im Anhang, Seite A1.

Als sekundäre Parameter wurden folgende Werte berechnet:

- (1) 26 weitere Maße und Indices, die sich aus den gemessenen Werten errechnen (K28 bis K53; siehe Anhang, Seite A2)
- (2) Werte einer die Geschlechter trennenden Diskriminanzfunktion (K54), in die nur 8 Formmerkmale eingehen; diese Diskriminanzfunktion wurde aus der Stichprobe selbst berechnet (siehe Kapitel 4.2.5., Seite 89).
- (3) 2 Faktoren aus einer Hauptkomponentenanalyse, in die alle 27 gemessenen Merkmale eingehen (K55 und K56)

4.1.3. Psychologische Daten

Da psychologische Aspekte nicht ganz unberücksichtigt bleiben sollen, wurde zur schnellen Erfassung von Persönlichkeitseigenschaften ein in der Psychologie weit verbreiteter Fragebogen verwendet. Der Fragebogen EPI ("Eysenck Personality Inventory", Eggert 1974) soll die Persönlichkeitsdimensionen "Extrovertiertheit" und "neurotische Tendenz" wiedergeben. Er kann nur sehr grobe Anhaltspunkte liefern, da er nur aus 57 ja/nein-Fragen besteht. Da er auch eine "Lügenskala" beinhaltet, kann abgeschätzt werden, ob die Versuchspersonen kooperativ waren.

Für die Auswertung des Tests wurde ein kleines Computerprogramm geschrieben, das die Skalenpunkte entsprechend der mitgelieferten Schablone registriert.

Von jeder Versuchsperson wurden die Punkte bezüglich Extrovertiertheit (PS1) und Neurotizismus (PS2) zu den übrigen Daten dazu abgespeichert.

In der untersuchten Stichprobe hatte niemand mehr als drei Punkte auf der Lügenskala, so daß alle Fragebögen verwendet werden konnten.

4.1.4. Auditive Beurteilungsdaten

4.1.4.1. Stimm – Selbstbeurteilung

Unmittelbar nach der Tonaufzeichnung der Sprechprobe bekam jede Versuchsperson ihre eigene Aufnahme über Kopfhörer vorgespielt und dazu einen Fragebogen (siehe Anhang, Seite A3) zum Ausfüllen, auf dem die Ausprägung (Skalen von 1 bis 7) von 25 Eigenschaften angekreuzt werden sollte. Der Fragebogen wurde weitgehend aus der Dissertation von Schneewind (1965) entnommen. Die 25 Skalenwerte (S1 bis S25) wurden zu den übrigen Daten in den Computer eingegeben.

4.1.4.2. Stimm – und Sprecher – Fremdbeurteilung

Sämtliche Sprechproben wurden fünf Beurteilern (Studenten der Phonetik, 4 weiblich, 1 männlich, Alter 20 – 30 Jahre) vorgespielt, die dann jeweils zwei Fragebögen auszufüllen hatten.

Der erste Fragebogen beinhaltet eine Stimm – und Sprechweisen – Beurteilung und erfaßt im wesentlichen dieselben 25 Attribute (B1 bis B25) wie der Stimm –

Selbstbeurteilungsfragebogen. (Einzelne Begriffe wurden durch andere ersetzt, weil sie sich mehr auf den semantischen Gehalt als auf die Stimme selbst beziehen und unter den standardisierten Versuchsbedingungen kaum sprecherspezifische Aussagen zulassen, siehe Anhang, Seite A4.)

Der zweite Fragebogen beinhaltet Eigenschaften, die sich auf die Person des Sprechers beziehen (siehe Anhang, Seite A5) und soll als Entsprechung zu den anthropometrischen und psychologischen Parametern den subjektiven Eindruck wiedergeben, den ein Sprecher im Zuhörer erweckt.

In beiden Fragebögen ist bei jedem Item zusätzlich eine Skala, die die subjektive Sicherheit des Beurteilers wiedergeben soll.

Da jedoch auf diesen Skalen zum Teil sehr stereotyp angekreuzt wurde (auf vielen Fragebögen wurde bei allen Items z. B. "5" angekreuzt), wurden sie nicht weiter ausgewertet.

Die Skalenwerte der Stimm- und Sprecherattribute wurden nun für jeden Beurteiler auf den Mittelwert = 0 und die Standardabweichung = 1 normiert (Z-Transformation) und dann über die fünf Beurteiler gemittelt und abgespeichert (B1 bis B44).

4.2. Geschlechtsdimorphismus

Bei vielen anthropologischen Merkmalssystemen läßt sich ein Großteil der interindividuellen Varianz auf den Geschlechtsdimorphismus zurückführen. Neben den mechanisch – funktionalen Aspekten (z. B. Geburtsvorgang) hat dieser Dimorphismus auch eine Signalwirkung, deren Funktion vor allem im Bereich des Sozialverhaltens liegt. Auch im Tierreich gibt es hierfür zahlreiche Beispiele.

An der Sprechstimme erkennt man meist sofort, ob ein Mann oder eine Frau spricht. In welcher Form diese Information in akustischen Parametern des Sprachsignals steckt, soll hier untersucht werden.

Zunächst werden alle akustischen Parameter getrennt auf ihre Geschlechtertrennfähigkeit hin untersucht.

Dann wird mit multivariaten Ähnlichkeitsvergleichen gezeigt, in welchen Spektren sich eine natürliche Gruppierung nach Geschlecht beobachten läßt.

Schließlich wird mit Hilfe der Cluster – und der Diskriminanzanalyse abgeschätzt, wie gut eine automatische Geschlechtsidentifizierung bei Verwendung von verschiedenen Parametersätzen funktioniert.

4.2.1. Varianzanalyse

Jeder der 276 akustischen Parameter wurde getrennt auf seinen Geschlechtsdimorphismus hin untersucht. Als Maß für die Trennschärfe wurde die Größe Wilk's Lambda (W) verwendet. Da sich in ihr aber nicht zeigt, ob im jeweiligen Parameter die Männer oder die Frauen im Schnitt den höheren Wert haben, wurde zusätzlich auch das Vorzeichen der Mittelwertsdifferenz

$$D = M(\text{männlich}) - M(\text{weiblich})$$

berücksichtigt. In Abbildung 27 ist folgendes Maß für jeden Parameter aufgezeichnet:

$$T = \sqrt{1 - W} \cdot \text{sign}(D)$$

Die Quadratwurzel wurde eingefügt, damit im Bereich der Signifikanzgrenze die graphische Auflösung besser ist.

Die gestrichelten Linien geben die Signifikanzgrenze für $P < 0.001$ an, die punktierten

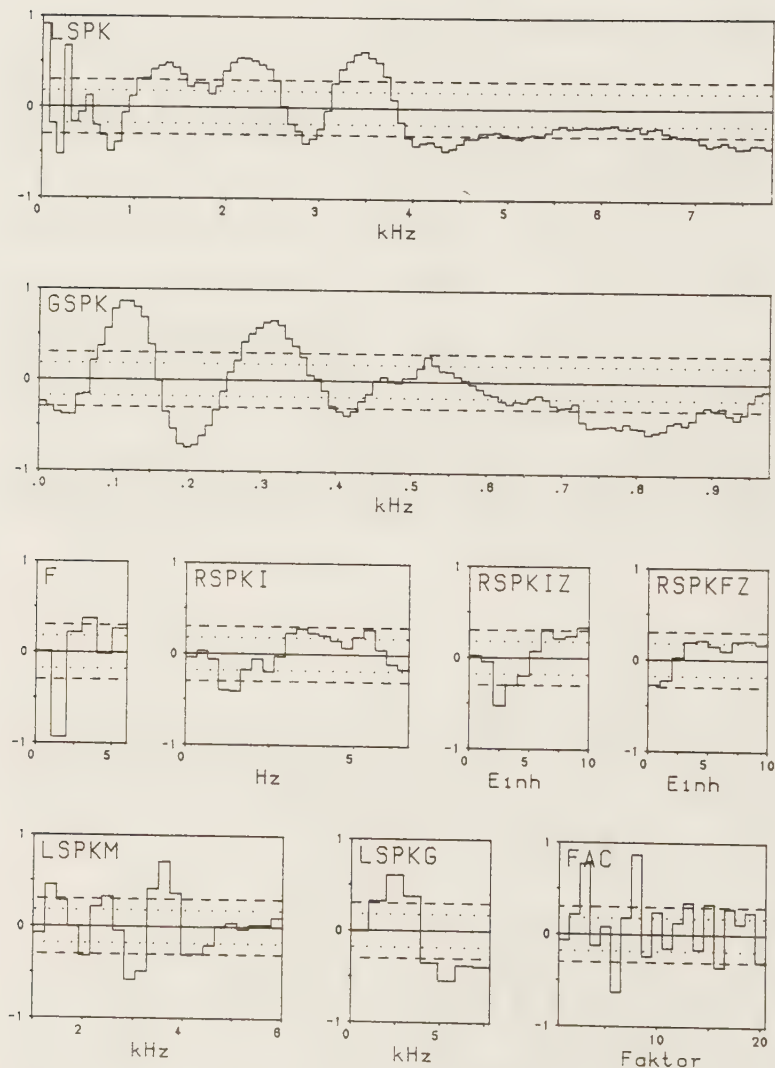


Abbildung 27:

Geschlechtsdimorphismus in 276 Parametern der menschlichen Sprechstimme:

Das Vorzeichen entspricht $\text{sign} (M(\text{männlich}) - M(\text{weiblich}))$;

Signifikanzgrenzen: gestrichelt $p < 0.001$, punktiert $p < 0.05$.

Linien die Signifikanzgrenze für $P < 0.05$.

Folgende Schlüsse lassen sich aus Abbildung 27 ziehen:

(1) Am stärksten ausgeprägt ist der Geschlechtsdimorphismus im Bereich der mittleren Grundfrequenz. Sie liegt bei Männern um ca. 100 Hz und bei Frauen um ca. 200 Hz, was am Grundfrequenzspektrum (GSPK) zu sehen ist. Der trennschärfste Einzelparameter ist die Grundfrequenz selbst (F2). Die erste aus dem Grundfrequenzspektrum extrahierte Hauptkomponente (FAC8) zeigt, daß ein großer Anteil der Varianz in diesem Frequenzbereich auf den Geschlechtsdimorphismus zurückgeführt werden kann.

(2) Auch in höheren Frequenzbereichen des Langzeitspektrums zeigen sich ausgeprägte geschlechtsspezifische Muster. Insbesondere haben Männer im Bereich um 3500 Hz eine höhere Intensität als Frauen. In den Teilstrukturen des Langzeitspektrums (LSPKM und LSPKG) und den aus ihnen extrahierten Faktoren (FAC1 bis FAC7) finden sich noch trennschärfere Einzelparameter als im Spektrum selbst.

(3) Auch im Sprechrhythmus gibt es signifikante Geschlechterunterschiede. Nachdem sich in der Sprechgeschwindigkeit (F1) keine Unterschiede zeigen, ist es verständlich, daß der Geschlechtsdimorphismus im auf die Lesezeit bezogenen Sprechrhythmus (der lineare Einfluß der Sprechgeschwindigkeit ist eliminiert) (RSPKIZ) stärker zum Ausdruck kommt als im absoluten Sprechrhythmus (RSPKI). Im Sprechrhythmus des Grundfrequenzverlaufs zeigen sich kaum Geschlechterunterschiede, lediglich in der Höhe der relativen Grundfrequenzschankungen (F3 bis F6).

4.2.2. Multivariate Ähnlichkeitsvergleiche

Um zu sehen, inwieweit sich bei multivariaten Ähnlichkeitsvergleichen mit verschiedenen Parametersätzen von selbst eine Gruppierung nach Geschlecht ergibt, wurden mit Hilfe der "Smallest Space Analysis" (Toledano 1977) für jede Sprechprobe Koordinaten in einem zweidimensionalen Raum berechnet und graphisch dargestellt (analog Abbildung 22, Seite 61); als Ähnlichkeitsmaß diente der Korrelationskoeffizient.

Da mit dem Verfahren aus technischen Gründen nur Stichproben bis zu 60 Individuen bearbeitet werden können, wurden jeweils zwei Analysen mit 58 (bzw. 59) Sprechproben durchgeführt.

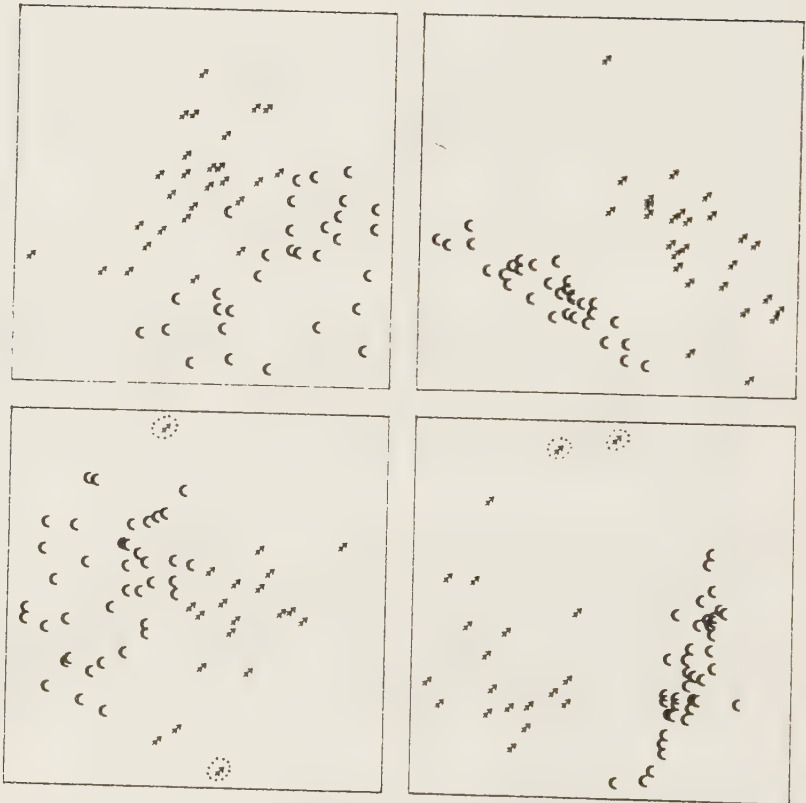


Abbildung 28:

Geschlechtsdimorphismus im multivariaten Ähnlichkeitsvergleich mit Hilfe der smallest space analysis:

oben: erster Teil der Stichprobe (Sprechprobe 1 bis 58)

unten: zweiter Teil der Stichprobe (Sprechprobe 59 bis 117)

links: Ähnlichkeit im Langzeitspektrum (LSPK, 100 Werte von 80 bis 8000 Hz)

rechts: Ähnlichkeit im Grundfrequenzspektrum (GSPK, 50 Werte von 10 bis 500 Hz)

Pfeil = männlich, Mond = weiblich,

punktiert eingekreist: Individuen, deren Geschlecht auditiv nicht erkannt wurde.

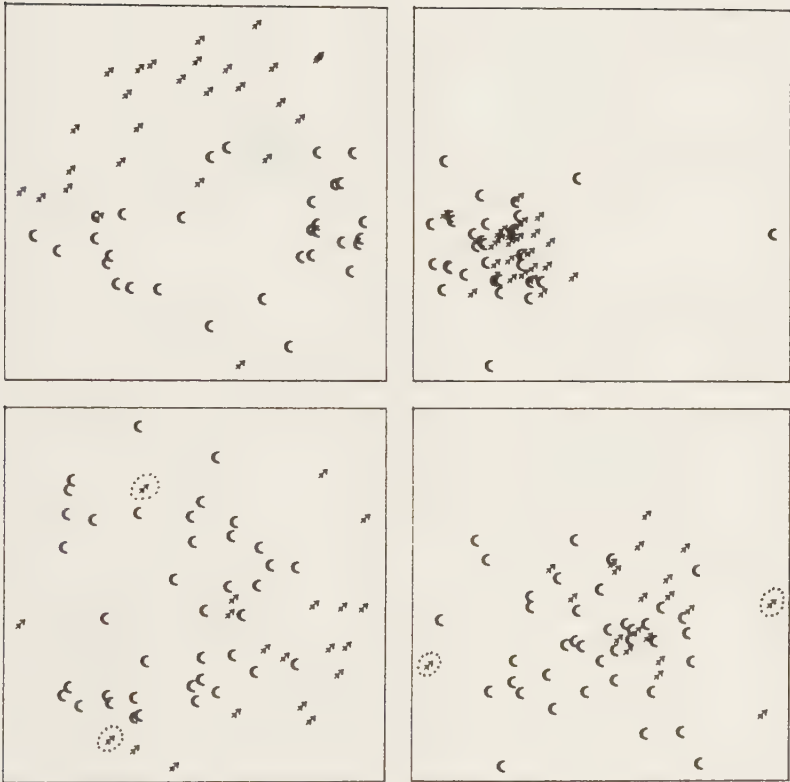


Abbildung 29:

Geschlechtsdimorphismus im multivariaten Ähnlichkeitsvergleich mit Hilfe der smallest space analysis:

oben: erster Teil der Stichprobe (Sprechprobe 1 bis 58)

unten: zweiter Teil der Stichprobe (Sprechprobe 59 bis 117)

links: Ähnlichkeit in der mittleren Struktur des Langzeitspektrums (LSPKM, 22 Werte von 1000 bis 6000 Hz)

rechts: Ähnlichkeit in der Grobstruktur des Langzeitspektrums (LSPKG, 8 Werte von 80 bis 8000 Hz)

Pfeil = männlich, Mond = weiblich,

punktiert eingekreist: Individuen, deren Geschlecht auditiv nicht erkannt wurde.

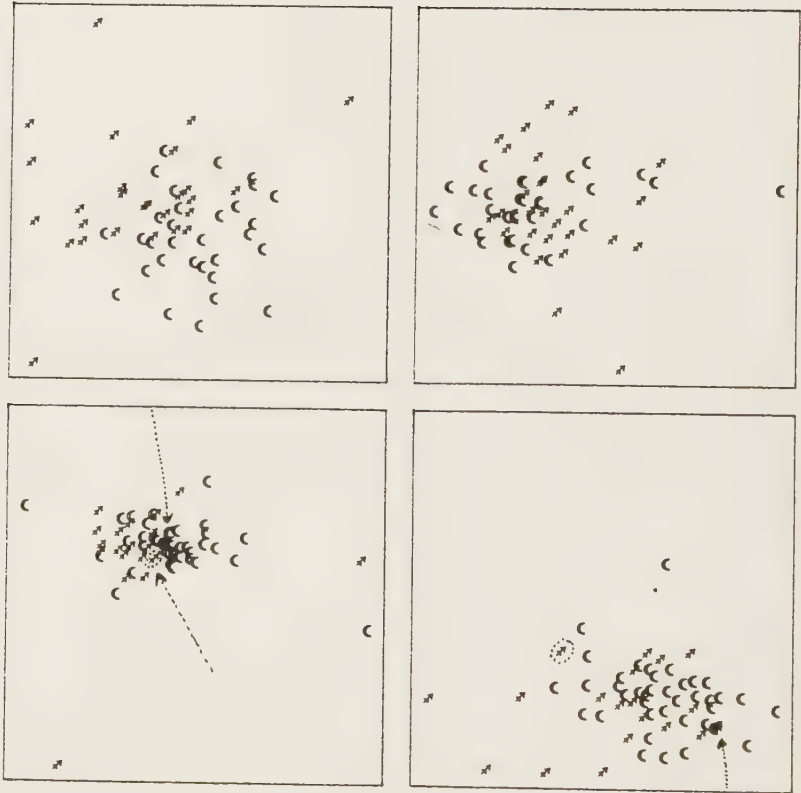


Abbildung 30:

Geschlechtsdimorphismus im multivariaten Ähnlichkeitsvergleich mit Hilfe der smallest space analysis:

oben: erster Teil der Stichprobe (Sprechprobe 1 bis 58)

unten: zweiter Teil der Stichprobe (Sprechprobe 59 bis 117)

links: Ähnlichkeit im Rhythmusspektrum des zeitkomprimierten Intensitätsverlaufs (RSPKIZ, 10 Werte)

rechts: Ähnlichkeit im Rhythmusspektrum des zeitkomprimierten Grundfrequenzverlaufs (RSPKFZ, 10 Werte)

Pfeil = männlich, Mond = weiblich,

punktiert eingekreist (bzw. mit punktiertem Pfeil markiert): Individuen, deren Geschlecht auditiv nicht erkannt wurde.

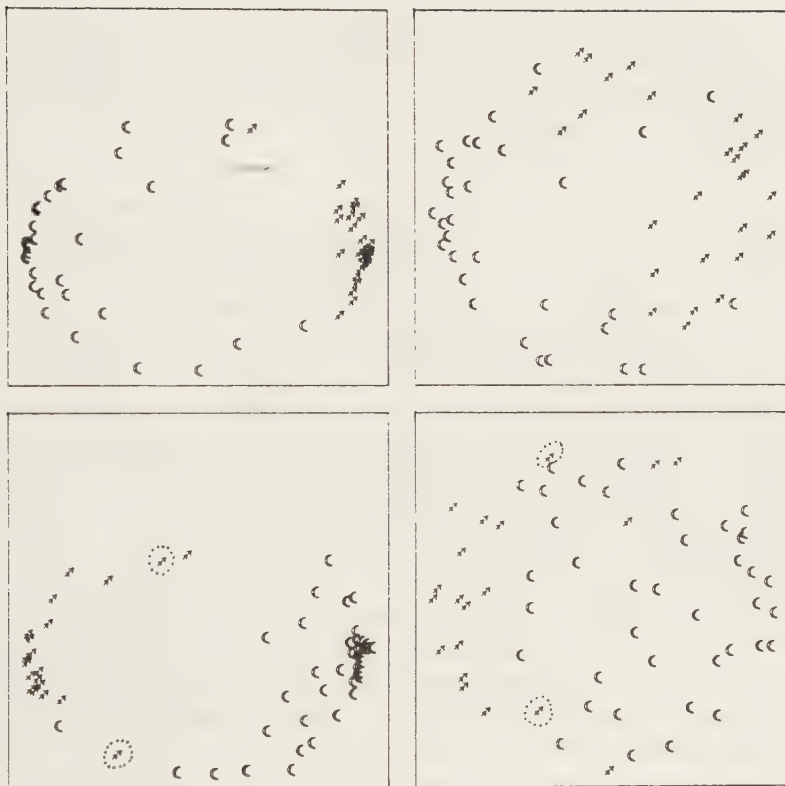


Abbildung 31:

Geschlechtsdimorphismus im multivariaten Ähnlichkeitsvergleich mit Hilfe der smallest space analysis:

oben: erster Teil der Stichprobe (Sprechprobe 1 bis 58)

unten: zweiter Teil der Stichprobe (Sprechprobe 59 bis 117)

links: Ähnlichkeit in 14 anthropometrischen Größenmerkmalen

rechts: Ähnlichkeit in 16 anthropometrischen Formmerkmalen

Pfeil = männlich, Mond = weiblich,

punktiert eingekreist: Individuen, deren Geschlecht auditiv nicht erkannt wurde.

Folgende Parametersätze wurden verwendet:

- (1) LSPK, 100 Werte von 80 bis 8000 Hz (Abbildung 28 links)
- (2) GSPK, 50 Werte von 10 bis 1000 Hz (Abbildung 28 rechts)
- (3) LSPKM, 22 Werte von 1000 bis 6000 Hz (Abbildung 29 links)
- (4) LSPKG, 8 Werte von 80 bis 8000 Hz (Abbildung 29 rechts)
- (5) RSPKIZ, 10 Werte (Abbildung 30 links)
- (6) RSPKIZ, 10 Werte (Abbildung 30 rechts)

Die beste Trennung läßt sich, wie zu erwarten ist, am Grundfrequenzspektrum beobachten. Beim Langzeitspektrum gibt es geringfügige Überlappungen; Ähnlich gut trennt die mittlere Struktur des Langzeitspektrums; Die Grobstruktur zeigt keine so gute Trennung. (Dies spricht wieder dafür, daß die Zerlegung in Teilstrukturen sinnvoll ist.) In den Rhythmusspektren sind leichte Tendenzen einer Geschlechtertrennung zu sehen.

Zum Vergleich mit der Geschlechtertrennfähigkeit anthropometrischer Merkmale wurde dasselbe Verfahren auch mit 14 Größenmerkmalen (Absolutmaße) und mit 16 Formmerkmalen (Indices) durchgeführt (Abbildung 31).

Folgende Maße wurden verwendet (siehe Anhang, Seite A1 und A2):

Größenmerkmale: K1, K2, K4, K5, K10, K15, K16, K17, K18, K20, K22, K23, K24, K27.

Formmerkmale: K35 bis K43 und K45 bis K51.

Alle Werte wurden vor der Analyse auf den Mittelwert = 0 und die Standardabweichung = 1 normiert, damit sie skalierungsunabhängig werden. Damit bei der Verwendung des Korrelationskoeffizienten als Ähnlichkeitsmaß auch Geschlechterunterschiede berücksichtigt werden, die bei allen Merkmalen dieselbe Tendenz aufweisen, wurde bei jedem zweiten Merkmal das Vorzeichen verändert. Wie Abbildung 31 zeigt, trennen Größenmerkmale besser als Formmerkmale; erstere trennen ungefähr so gut wie die Werte des Grundfrequenzspektrums, zweitere so gut wie die Parameter aus der mittleren Struktur des Langzeitspektrums oberhalb 1000 Hz.

4.2.3. Automatische Geschlechtererkennung

Für eine automatische Geschlechtererkennung gibt es zwei Möglichkeiten:

- (1) Bei einem sehr ausgeprägten Dimorphismus ergibt sich mit Hilfe der Clusteranalyse eine Trennung so, daß die einzelnen Cluster in Bezug auf das Geschlecht homogen sind. Man braucht von jedem Cluster nur ein einziges Individuum bekannten Geschlechts und erhält eine vollständige Zuordnung.

(2) Man hat eine Lernstichprobe und führt mit ihr eine Diskriminanzanalyse durch; unbekannte Proben können dann mit Hilfe der Diskriminanzfunktion zugeordnet werden.

Die Clusteranalyse hat den Vorteil, daß man auch bei kleinen Stichproben viele Merkmale eingehen lassen kann und nur wenige Individuen mit bekanntem Geschlecht benötigt; die Diskriminanzanalyse erzielt bei Verwendung weniger Merkmale oft schon gute Ergebnisse und ermöglicht besser, den zu erwartenden Zuordnungsfehler durch Signifikanztests abzuschätzen.

4.2.3.1 Clusteranalyse

Die Clusteranalyse wurde mit der Prozedur "QUICK CLUSTER" des SPSSX-Programmpakets (Schubö und Üehlinger 1984) durchgeführt. Aus der Gesamtheit der Sprechproben wurden jeweils sechs Cluster gebildet. (Für die Darstellung eines Dimorphismus sollten eigentlich zwei Cluster genügen; es hat sich aber herausgestellt, daß es Individuen gibt, die allen anderen sehr unähnlich sind, so daß Cluster vorkommen, die nur ein einziges Individuum beinhalten.)

Die Analyse wurde mit folgenden Datensätzen durchgeführt:

- (1) LSPK (100 Werte)
- (2) GSPK (50 Werte)
- (3) LSPKG und LSPKM gemeinsam (30 Werte)
- (4) RSPKI und RSPKIZ gemeinsam (30 Werte)
- (5) FAC (20 Werte)

Inwieweit die Cluster zwischen den Geschlechtern trennen, geht aus Tabelle 1 hervor. Aufgrund der Zusammensetzung der Gesamtheit von 47 Männern und 70 Frauen ergibt sich eine a priori- Erwartung für das Geschlechterverhältnis in den Clustern von 47/70. Je stärker die tatsächlichen Werte davon abweichen, desto ausgeprägter ist der Dimorphismus.

Bei allen Analysen finden sich Cluster mit markanten Abweichungen von der Zufallserwartung; als zuverlässig geschlechtertrennend können allerdings nur die nach dem Grundfrequenzspektrum erstellten Cluster angesehen werden.

Cluster	Sex	LSPK	GSPK	LSPKM	RSPKIZ	FAC
				LSPKG	RSPKFZ	
1	M	14 *	30	2	14	7 **
	W	41	-	9	9	64
2	M	17	1	1	2 *	38
	W	-	-	-	24	2
3	M	-	1 *	26	1	1
	W	4	24	-	3	-
4	M	-	-	5	5	2
	W	25	10	13	3	1
5	M	1	-	1 *	13 *	-
	W	-	36	22	25	1
6	M	15 *	15 *	12 *	12	1
	W	-	-	26	6	-

Tabelle 1:

Geschlechtertrennung mit Hilfe der Clusteranalyse:

Mit der Prozedur "QUICK CLUSTER" (SPSS) wurden für jeden der fünf aufgeführten Parametersätze aus der Sprechstimme sechs Cluster erzeugt; die Zufallserwartung für das Geschlechterverhältnis beträgt 47/70 ; starke Abweichungen davon können dem Geschlechtsdimorphismus im jeweiligen Parametersatz zugeschrieben werden.

*In den mit * markierten Clustern befindet sich ein Individuum, dessen Geschlecht auditiv nicht erkannt wurde.*

4.2.3.2. Diskriminanzanalyse

Mit Hilfe der Prozedur "DISCRIMINANT" (SPSS) wurde die Geschlechtertrennung empirisch getestet:

Es wurden mit jedem Parametersatz zwei Analysen gerechnet; bei jeder wurde eine Zufallsstichprobe von 20 männlichen und 30 weiblichen Proben weggelassen (jeweils verschiedene Individuen), mit denen dann der Prozentsatz richtiger Zuordnungen ermittelt werden konnte.

Mit folgenden Parametersätzen wurden Diskriminanzanalysen durchgeführt:

- (1) FAC 1 bis 20 (Rhythmus und Klang zusammen)
- (2) FAC 1 bis 11 (nur Klangparameter)
- (3) FAC 1 bis 7 (nur Teilstrukturen des Langzeitspektrums oberhalb 1000 Hz)
- (4) FAC 1 bis 5 (nur aus der mittleren Struktur des Langzeitspektrums)
- (5) LSPKM 8 bis 13 (mittlere Struktur des Langzeitspektrums von ca. 2800 bis 4000 Hz)
- (6) LSPKG 1 bis 8 (Grobstruktur des Langzeitspektrums)
- (7) RSPKIZ 1 bis 10 (Rhythmusspektrum des zeitkomprimierten Intensitätsverlaufs)
- (8) FAC 12 bis 18 (alle Rhythmusfaktoren)

Zum Vergleich wurde auch die Geschlechtertrennung mit anthropometrischen Merkmalen durchgeführt:

- (9) 8 Größenmerkmale (K1, K4, K10, K15, K16, K22, K23, K24)
- (10) 8 Formmerkmale (K35, K36, K37, K41, K45, K46, K47, K49)

Sämtliche aufgestellten Diskriminanzfunktionen ergaben hochsignifikante Trennungen; selbst der am schlechtesten trennende Parametersatz (FAC 12–18) ist noch signifikant ($P < 0.002$).

Bemerkenswert ist insbesondere, daß der Geschlechtsdimorphismus auch im Sprechrhythmus zum Ausdruck kommt.

Parameter aus der Sprechstimme:

Parameter- satz	Werte	kanonische Korrelation	richtige Zuordnung unbekannter Proben
FAC	1-20	0.97	99 % *
FAC	1-11	0.97	99 % *
FAC	1- 7	0.87	94 % *
FAC	1- 5	0.83	94 % *
LSPKM	8-13	0.78	85 % **
LSPKG	1- 8	0.73	82 % *
RSPKIZ	1-10	0.64	73 % *
FAC	12-18	0.57	66 % *

Anthropometrische Merkmale:

Parameter- satz	Werte	kanonische Korrelation	richtige Zuordnung unbekannter Proben
Size	8	0.91	98 % *
Shape	8	0.87	96 %

Tabelle 2: Geschlechtertrennung mit Hilfe der Diskriminanzanalyse nach Parametern aus der Sprechstimme und nach anthropometrischen Merkmalen:

Size: folgende 8 Größenmerkmale: K1, K4, K10, K15, K15, K22, K23, K42;

Shape: folgende 8 Formmerkmale: K35, K36, K37, K41, K45, K46, K47, K49;

*Bei den mit * markierten Analysen wurde ein (** = zwei) Individuum, dessen Geschlecht auditiv nicht erkannt wurde, auch durch die Diskriminanzanalyse falsch klassifiziert.*

4.2.4. Auditive Geschlechtswahrnehmung

Bei der Durchführung der auditiven Tests mit fünf Beurteilern, denen alle 117 Sprechproben vorgespielt wurden, wurde bei einer Sprechprobe von allen Beurteilern das Geschlecht nicht erkannt, bei einer weiteren Sprechprobe erkannten nur 2 der Beurteiler das Geschlecht.

Die auditive Erkennungsrate liegt also zwischen 98 und 99% und ist somit nicht höher als die nach objektiven akustischen Parametern.

Möchte man ganz allgemein etwas über die Arbeitsweise von Erkennungssystemen wissen, so bietet die Analyse ihrer Fehler oft gute Anhaltspunkte. Machen zwei Systeme dieselben Fehler, so liegt nahe, daß sie auch ähnlich arbeiten.

Angewendet auf die Geschlechtsbestimmung heißt dies: Mit welchen Parametersätzen macht die objektive Geschlechterklassifizierung dieselben Fehler wie die auditive? Zur Beantwortung dieser Frage wurden in den Ähnlichkeitsdiagrammen die beiden auditiv falsch geschlechtsklassifizierten Proben gepunktet eingekreist (bzw. mit einem gepunkteten Pfeil markiert).

(Siehe Abbildung 28 – 31, Seite 78 – 81)

Sie liegen oft außerhalb der männlichen Variabilität, aber nicht immer in Richtung der weiblichen.

LSPK: Einer der beiden Sprecher würde als Frau eingeordnet, der andere als Mann (Abbildung 28 links unten).

GSPK: Beide liegen deutlich außerhalb der weiblichen Variabilität, allerdings auch nicht richtig innerhalb der männlichen (Abbildung 28 rechts unten).

LSPKM: Beide liegen mehr im weiblichen als im männlichen Streuungsbereich (Abbildung 29 links unten).

LSPKG: Eine erscheint extrem weiblich, die andere extrem männlich (Abbildung 29 rechts unten).

RSPKIZ: Beide liegen an der Grenze, aber eher im weiblichen als im männlichen Bereich (Abbildung 30 links unten).

RSPKFZ: Eine der beiden Proben liegt mehr im weiblichen Streuungsbereich, die andere mehr im männlichen (Abbildung 30 rechts unten).

Aufgrund der Tatsache, daß in keinem Parametersatz beide Proben innerhalb der männlichen Variabilität liegen, ist es offensichtlich ein sehr komplexes Bündel von Einzelphänomenen, das diese Stimmen weiblich erscheinen läßt. Was im einzelnen für die Zuhörer den Ausschlag bei der Geschlechtszuordnung gegeben hat, kann daher hier

nicht entschieden werden.

Weiterhin wurde nun untersucht, wo sich diese beiden Sprecher in den Ähnlichkeitsdiagrammen finden, die nach anthropometrischen Merkmalen erstellt wurden. Wie Abbildung 31 zeigt, liegen beide sowohl in dem nach Größenmerkmalen als auch in dem nach Formmerkmalen aufgestellten Diagramm im Grenzbereich zwischen der männlichen und der weiblichen Variabilität.

Obwohl es sich hier um eine Beobachtung an nur zwei Individuen handelt, ist es doch unwahrscheinlich, daß diese Parallele zwischen Stimme und Körper rein zufällig zustandegekommen ist.

Wie sich die beiden Sprechproben in den Ergebnissen der Cluster- und der Diskriminanzanalyse verhalten, geht aus Tabelle 1 (Seite 84) und Tabelle 2 (Seite 86) hervor:

In der Clusteranalyse werden sie in der Mehrzahl der Fälle Clustern zugeordnet, die wesentlich mehr Frauen als Männer beinhalten.

In der Diskriminanzanalyse befindet sich bei jedem eingehenden Parametersatz mindestens einer der beiden Sprecher unter den falsch eingeordneten Individuen. Insbesondere wurde auch bei Verwendung anthropometrischer Merkmale einer der beiden Sprecher als Frau eingeordnet.

4.2.5. Sprechstimme und Andromorphie innerhalb eines Geschlechts

Die zwei Einzelbeobachtungen, an den auditiv in Bezug auf das Geschlecht nicht erkannten Individuen, die auch in anthropometrischen Merkmalen im Grenzbereich zwischen den Geschlechtern liegen, könnten den Schluß nahelegen, es bestehe ein allgemeiner Zusammenhang zwischen Maskulinität in der Stimme und Andromorphie auch innerhalb der Geschlechtergruppen.

Um dies zu testen, wurde zunächst mit allen 117 Individuen unter Verwendung von 8 Formmerkmalen (K35, K36, K37, K41, K45, K46, K47, K49) eine Diskriminanzanalyse in Bezug auf das Geschlecht durchgeführt; die Diskriminanzwerte jedes Individuums wurden abgespeichert. Da sie bei Männern die höheren Werte aufweisen, seien sie als Andromorphie – Scores (K54) bezeichnet.

Für alle 276 Parameter aus der Sprechstimme wurde nun, getrennt bei beiden Geschlechtern, die Korrelation mit K54 berechnet und graphisch dargestellt. Die Hauptkomponenten aus den Spektren (FAC) wurden hierfür bei beiden Geschlechtern getrennt extrahiert.

Aus den Abbildungen 32 und 33 geht hervor, daß sich kein signifikanter Zusammenhang nachweisen läßt, daß also bezüglich der normalen Variabilität die Andromorphie innerhalb eines Geschlechts in den aus der Sprechstimme berechneten Parametern nicht zum Ausdruck kommt. Die gestrichelte Linie gibt die Signifikanzschwelle für $p < 0.01$ an; bei fast 300 Einzelparametern müßte sie von wesentlich mehr als drei Werten deutlich überschritten sein, damit sich das Ergebnis vom Zufall abgrenzen läßt.

Um dies zu demonstrieren, wurde statt K54 eine Zufallszahl mit den Stimmparametern korreliert; das resultierende Muster findet sich in Abbildung 34.

Offensichtlich stellt also der Geschlechtsdimorphismus eine distinkte Unterteilung der Gesamtheit dar; er kann nicht im Sinne eines Kontinuums verstanden werden, in dem die einzelnen Individuen sich ganzheitlich einordnen lassen.

Die in 4.2.4. beschriebene Beobachtung an den auditiv falsch eingeordneten Sprechproben läßt sich also nicht auf die Gesamtheit übertragen. Ursachen für die Sonderstellung der beiden Sprecher konnten hier nicht gefunden werden.

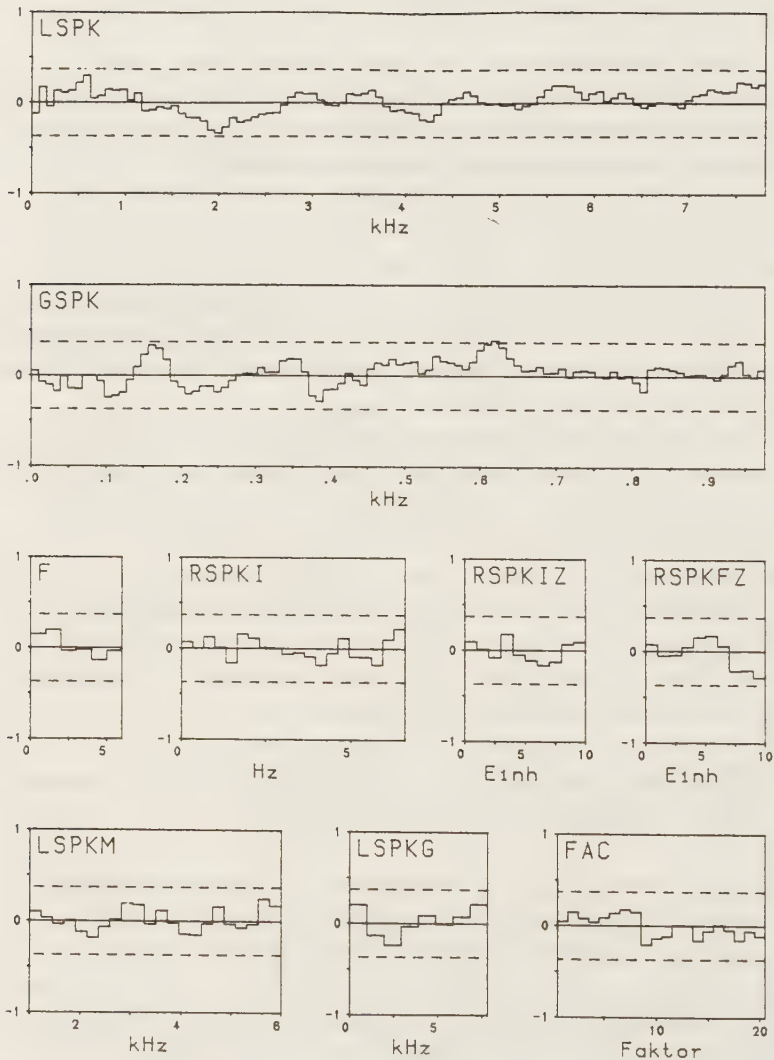


Abbildung 32:

Korrelation von Parametern aus der Sprechstimme mit den aus der Stichprobe mit Hilfe der Diskriminanzanalyse errechneten Andromorphie – Scores:

Männliche Stichprobe, $n=47$

gestrichelt: Signifikanzgrenze für $p < 0.01$

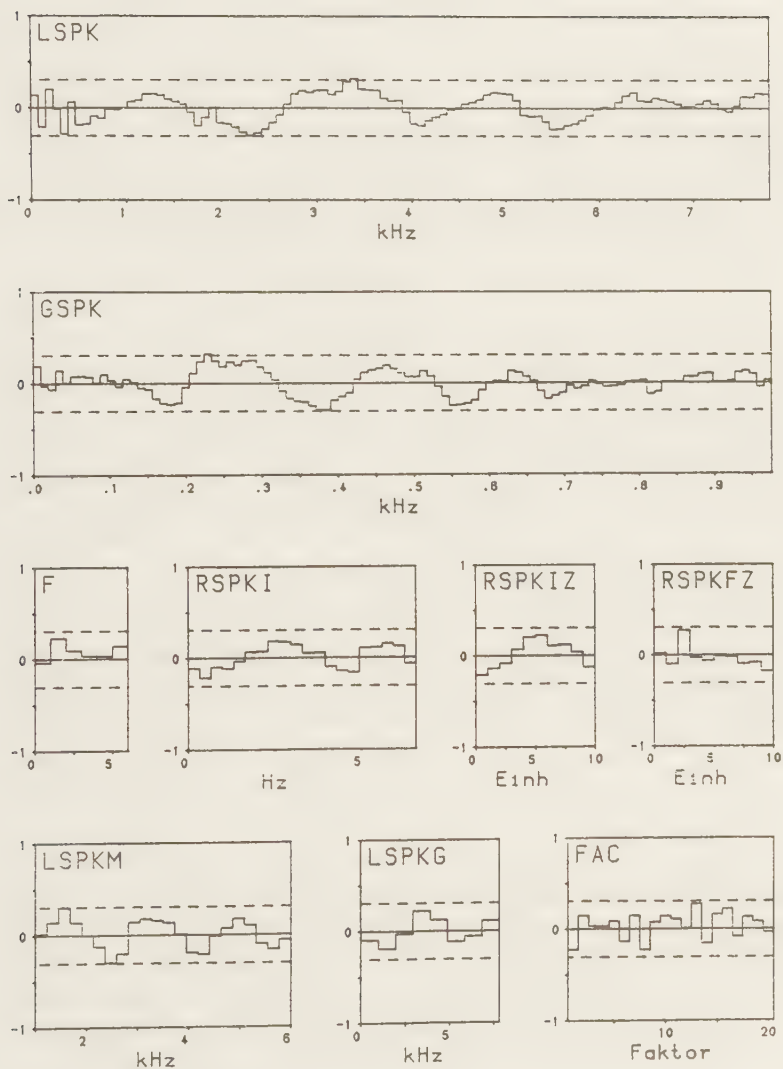


Abbildung 33:

Korrelation von Parametern aus der Sprechstimme mit den aus der Stichprobe mit Hilfe der Diskriminanzanalyse errechneten Andromorphie – Scores:

Weibliche Stichprobe, $n = 70$

gestrichelt: Signifikanzgrenze für $p < 0.01$

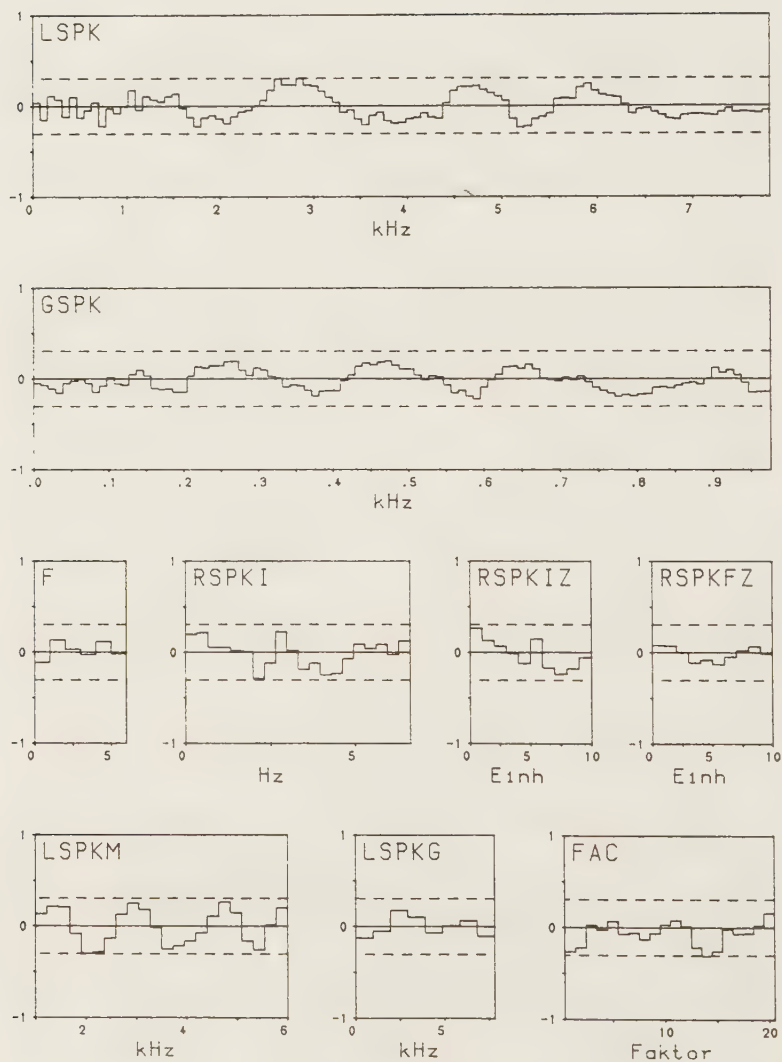


Abbildung 34:

Korrelation von Parametern aus der Sprechstimme mit einer Zufallszahl:

Weibliche Stichprobe, $n = 70$

gestrichelt: Signifikanzgrenze für $p < 0.01$

4.3. Auditive Stimmbeurteilung und akustische Korrelate

4.3.1. Selbst – und Fremdbeurteilung

Dieser Beitrag ist mehr methodisch zu verstehen: Es soll geprüft werden, inwieweit eine Versuchsperson in der Lage ist, ihre eigene Stimme nach Anhören über Kopfhörer zu beurteilen. Für den Fall, daß sich eine gute Übereinstimmung zwischen Selbst – und Fremdbeurteilung ergibt, könnte man sich in künftigen Untersuchungen eventuell auf die Selbstbeurteilung beschränken, was den Aufwand erheblich reduziert.

Zunächst wurde geprüft, ob eine Übereinstimmung besteht zwischen dem allgemeinen Sprachempfinden und dem aus den Fragebogendaten zu schließenden Verständnis der Stimmattribute von Seiten der Beurteiler und der Sprecher. Hierfür wurde die Korrelationsmatrix der Beurteilungsdaten mit Hilfe der Smallest Space Analysis in zwei Dimensionen dargestellt (Abbildung 35 und 36). Begriffe mit ähnlicher Bedeutung sollten nahe beieinander stehen, solche mit gegensätzlicher Bedeutung weiter voneinander entfernt.

Aus den Abbildungen ist ersichtlich, daß dies sowohl für die Fremd – als auch für die Selbstbeurteilung zutrifft; die Fragebögen wurden also nicht arbiträr sondern, zumindest was die Relation der Attribute zueinander betrifft, konsistent angekreuzt. Beide Abbildungen zeigen ähnliche Muster, wenn man eine davon um 180° dreht. Man könnte ganz grob sagen, daß die waagrechte Achse sich mehr auf den Redefluß (Zeitverlauf der Sprechprobe) bezieht, die senkrechte mehr auf den Klang.

Zur Überprüfung der Übereinstimmung von Fremd – und Selbstbeurteilung wurden die 20 beiden Fragebögen gemeinsamen Begriffe in einer Abbildung dargestellt (Abbildung 37). Es zeigt sich, daß die Relation der Begriffe zueinander innerhalb der beiden Datensätze besser übereinstimmt als dieselben Begriffe zwischen den Datensätzen; es bestehen also offensichtlich systematische Verschiebungen. Interessant ist, daß sich auch Begriffe, die mit einer negativen Wertung verbunden sind (z. B. "zaghaft"), in der Selbstbeurteilung durchaus in derselben Richtung absetzen wie in der Fremdbeurteilung.

Über die Zuverlässigkeit auditiver Beurteilungen und Selbstbeurteilungen gibt Tabelle 3 Auskunft, die nach dem Modell in Abbildung 38 zusammengestellt wurde:

Für jede Eigenschaft wurden paarweise unter den 5 Beurteilern die Interkorrelationen berechnet (10 Korrelationskoeffizienten); r_1 ist das arithmetische Mittel davon. Mit einem Stern wurden Werte versehen, bei denen mindestens 6 der 10 Einzel – Korrelations-

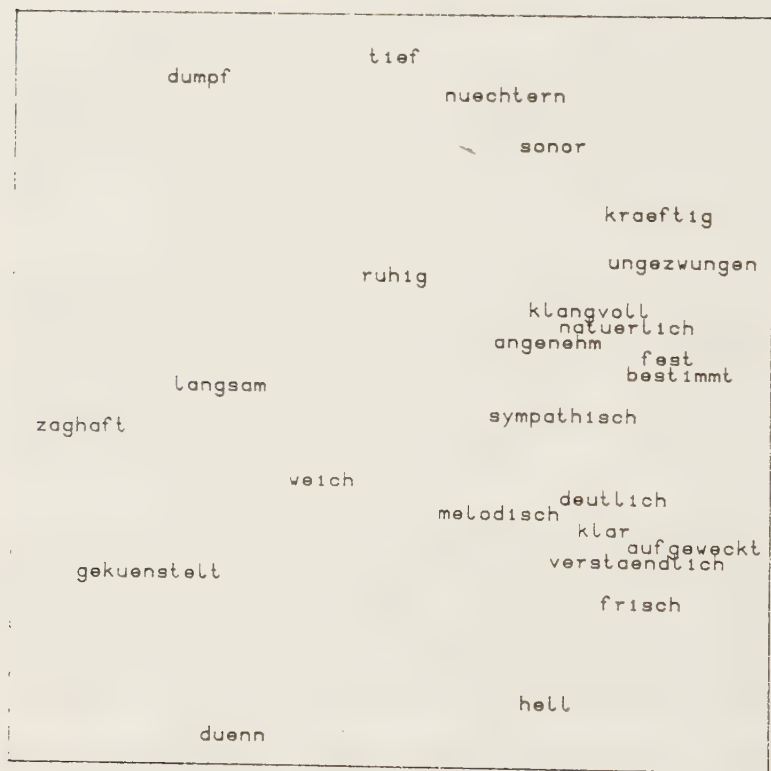


Abbildung 35:

Darstellung der Ähnlichkeit von Attributen der Sprecher – Fremdbeurteilung:

Die Korrelationsmatrix der Begriffe wurde mit Hilfe der "smallest space analysis" (Prozedur "SSA1" aus SPSS, Toledamo 1977) zweidimensional abgebildet. (Das Verfahren kann sowohl dafür verwendet werden, Merkmale abzubilden, als auch dafür, Merkmalsträger nach Ähnlichkeit zu gruppieren, wie in den vorangehenden Kapiteln.)

Es lassen sich in wesentlichen zwei Polarisierungen erkennen: Die Horizontale stellt hauptsächlich den Redefluß und die Sprechgeschwindigkeit dar (z. B. das Gegensatzpaar "zaghaft" – "fest") die Vertikale bezieht sich mehr auf den Klang und die Tonhöhe, (z. B. "tief" – "hell").

Für die Lokalisation der Begriffe ist der Anfangsbuchstabe maßgeblich.

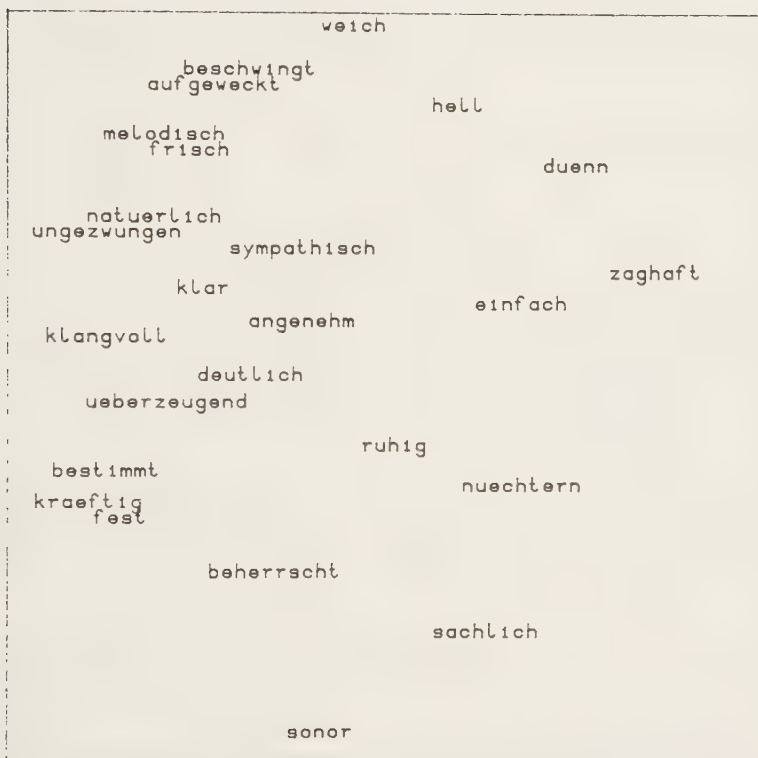


Abbildung 36:

Darstellung der Ähnlichkeit von Attributen der Sprecher – Selbstbeurteilung:

Die Korrelationsmatrix der Begriffe wurde mit Hilfe der "smallest space analysis" zweidimensional abgebildet (analog Abbildung 35). Es zeigt sich ein ähnliches Muster wie in der Fremdbeurteilung; die Aufspaltung der Gegensatzpaare ist weitgehend dieselbe.

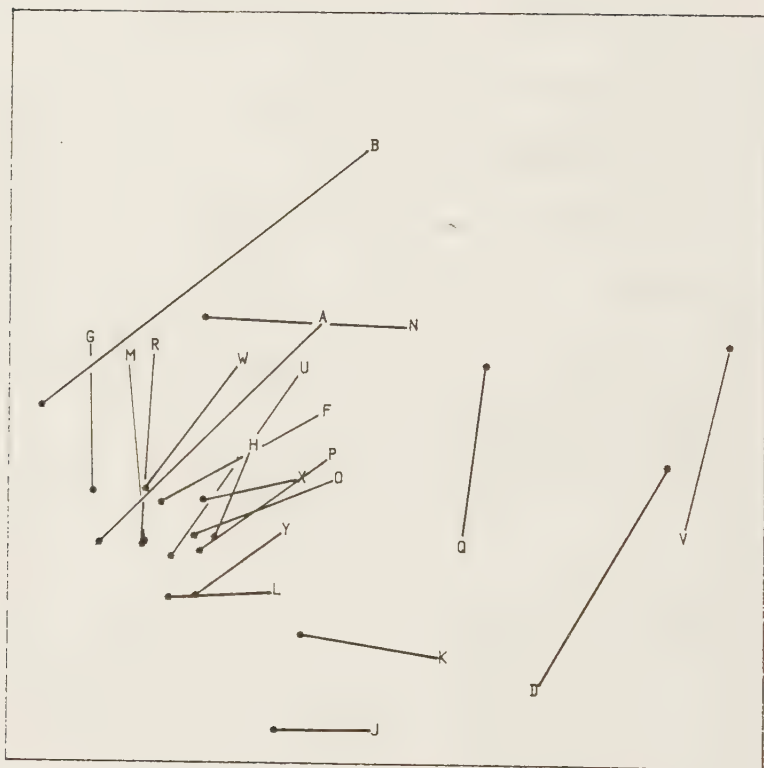


Abbildung 37:

Darstellung der Ähnlichkeit von Attributen der Sprecher- Selbstbeurteilung und Fremdbeurteilung:

Es wurden die 20 beiden Fragebögen gemeinsamen Begriffe verwendet.

Die Buchstaben bezeichnen die Selbstbeurteilungsattribute, die durch eine Linie mit ihnen verbundenen Sterne die entsprechenden Fremdbeurteilungsattribute.

(A = ungezwungen, B = sonor, D = dünn, F = angenehm, G = kräftig, H = deutlich, J = nüchtern, K = hell, L = aufgeweckt, M = bestimmt, N = ruhig, O = sympathisch, P = klar, Q = weich, R = fest, U = natürlich, V = zaghaft, W = klangvoll, X = melodisch, Y = frisch)

Die Tatsache, daß die Verbindungslinien trotz der zum Teil schlechten Übereinstimmung zwischen Selbst- und Fremdbeurteilung nicht zufällig im Raum liegen, zeigt, daß offensichtlich systematische Verschiebungen bestehen.

Koeffizienten für $p < 0.01$ signifikant waren, mit zwei Sternen die, bei denen mindestens 6 für $p < 0.001$ signifikant waren.

Unmittelbar mit r_1 kann r_2 verglichen werden; r_2 ist das arithmetische Mittel aus den 5 Korrelationskoeffizienten der Selbstbeurteilungs- mit den Beurteilerdaten. Sterne wurden vergeben, wenn mindestens 3 der Einzelwerte für $p < 0.01$ bzw. $p < 0.001$ signifikant waren.

Die Übereinstimmung der zusammengefaßten Beurteilerdaten (Mittelwerte über die z -transformierten Daten der 5 Beurteiler) mit der Selbstbeurteilung gibt r_3 wieder; p_3 ist das Signifikanzniveau von r_3 , in Zahlen ausgedrückt.

Folgende Aussagen lassen sich nach Tabelle 3 machen:

(1) Die zum Teil niedrigen Werte von r_1 zeigen, daß Urteile einzelner Zuhörer weit auseinandergehen können. Dies kann einerseits an einem unterschiedlichen Verständnis der Begriffe liegen, andererseits auch (besonders bei Begriffen, mit denen ein Werturteil impliziert ist) am persönlichen Geschmack.

(2) Daß r_2 im Schnitt niedriger ist als r_1 , kann entweder damit erklärt werden, daß die Gruppe der Selbstbeurteiler sehr heterogen ist und somit Zusammenhänge nur unschärfer herauskommen, oder damit, daß die Selbstbeurteiler ihrer eigenen Stimme gegenüber befangen sind und dadurch nicht das beurteilen, was sie über Kopfhörer hören, sondern das, was sie als Bild von sich selbst mit sich herumtragen.

Um zwischen diesen beiden Erklärungsangeboten entscheiden zu können, kann r_3 verwendet werden: Ist sowohl r_3 als auch r_2 niedrig, obwohl r_1 hoch ist, so kann man die Befangenheit der Selbstbeurteiler dafür verantwortlich machen. Dies ist vor allem bei den Begriffen "ungezwungen" und "natürlich" der Fall. Liegt r_2 in derselben Größenordnung wie r_1 , so heißt das, daß die Selbstbeurteilung einer Fremdbeurteilung äquivalent ist. Ist r_2 niedriger als r_1 , r_3 aber wesentlich höher als r_2 , so heißt das, daß die Selbstbeurteilung zwar nicht ganz so zuverlässig ist wie die Fremdbeurteilung, daß aber vor allem die interindividuelle Streuung dafür verantwortlich ist und nicht die Befangenheit der Selbstbeurteiler.

(3) Die Tatsache, daß r_3 oft höher ist als r_1 , auch wenn r_2 niedriger ist als r_1 , spricht dafür, daß die Mittelung über die 5 Beurteiler ein wesentlich zuverlässigeres Bild von der Sprechprobe liefert als die Daten der einzelnen Beurteiler. In den folgenden Kapiteln werden daher nur noch die von den 5 Beurteilern zusammengefaßten Daten verwendet.

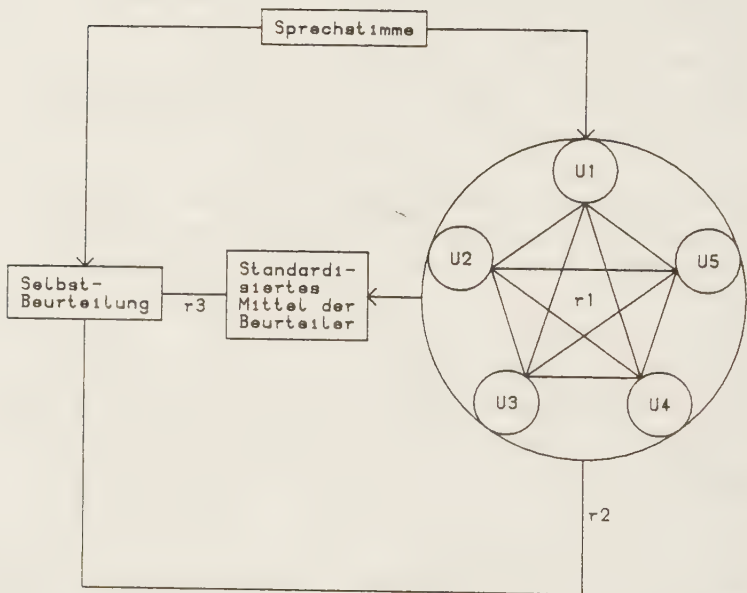


Abbildung 38:

Modell zum Vergleich der Selbst- und Fremdbeurteilung von Sprechproben:

Die Daten der Beurteiler (U_1, U_2, U_3, U_4, U_5) wurden interkorreliert (das Pentagon mit den Innenlinien symbolisiert die 10 möglichen Korrelationskoeffizienten). Das Mittel aus den 10 Werten ist r_1 .

Das Mittel aus den fünf Korrelationskoeffizienten der Selbstbeurteilung mit den Fremdbeurteilungen ist als r_2 bezeichnet.

r_3 ist der Korrelationskoeffizient zwischen Selbstbeurteilung und dem Mittelwert der standardisierten (bezüglich Mittelwert und Streuung) Beurteilerdaten.

Die konkreten Zahlenwerte für die einzelnen Attribute sind in Tabelle 3 aufgeführt; signifikante Werte sind markiert.

Nr.	Attribut	r1	r2	r3	p3
1	ungezwungen	.30 *	.06	.10	.150
2	sonor	.23 *	.24 *	.38 **	.000
3	tief	.33 **			
4	dünn	.25	.04	.07	.241
5	gekünstelt	.24 *			
6	angenehm	.21	.15	.25 *	.003
7	kräftig	.32 *	.28 *	.42 **	.000
8	deutlich	.30 *	.20	.30 *	.001
9	langsam	.46 **			
10	nüchtern	.11	.10	.19	.022
11	hell	.35 **	.19	.27 *	.001
12	aufgeweckt	.32 **	.22	.33 **	.000
13	bestimmt	.32 *	.17	.26 *	.003
14	ruhig	.20	.16	.27 *	.001
15	sympathisch	.19	.12	.21	.012
16	klar	.24	.19 *	.31 **	.000
17	weich	.19	.03	.05	.306
18	fest	.27 *	.14	.21	.011
19	dumpf	.16			
20	verständlich	.17			
21	natürlich	.26 *	.04	.06	.266
22	zaghaft	.34 **	.25 *	.36 **	.000
23	klangvoll	.17	.19 *	.33 **	.000
24	melodisch	.18	.19	.33 **	.000
25	frisch	.31 **	.21 *	.32 **	.000

Tabelle 3:

Korrelationen von Attributen zwischen Beurteilern (r1), zwischen Beurteilern und Selbstbeurteilung (r2) und zwischen den über 5 Beurteiler gemittelten Daten und der Selbstbeurteilung (r3).

p3 ist das Signifikanzniveau von r3; bei r1 und r2 wurden Sterne gedruckt, wenn mindestens 60% der eingehenden Korrelationskoeffizienten signifikant waren (*: $p < .01$; **: $p < .001$). Das der Tabelle zugrundeliegende Modell ist in Abbildung 38 dargestellt.

4.3.2. Auditive Stimmbeurteilung und akustische Parameter

4.3.2.1. Faktorenstruktur der Stimmbeurteilungsattribute

Nachdem sich unter den 25 Attributen des Stimmbeurteilungsfragebogens einige von ihrer Bedeutung her stark ähneln oder gegensätzlich sind, ist anzunehmen, daß in den Daten eine gewisse Redundanz steckt. Da bei der Anwendung multivariater statistischer Methoden möglichst wenige, aussagekräftige Variablen erwünscht sind, wurde zur Datenreduktion eine Hauptkomponentenanalyse durchgeführt. Auf eine spezielle Rotation des Faktorraums wurde verzichtet, da sich bereits die Hauptkomponenten interpretieren lassen. Für beide Geschlechter wurden getrennte Analysen durchgeführt, damit sich bei den nachfolgenden Vergleichsuntersuchungen bezüglich akustischer Korrelate nicht nur der (in der Stimme ja stark ausgeprägte) Geschlechtsdimorphismus abbildet.

Es wurden 4 Hauptkomponenten berechnet (FB1 bis FB4); mit ihnen wird ein Varianzanteil von 75% bei der männlichen und von 72% bei der weiblichen Stichprobe beschrieben. Über die Bedeutung der Faktoren gibt Tabelle 4 Auskunft; stark positive Ladungen wurden mit " + " gekennzeichnet, stark negative mit " - " und schwache (positiv oder negativ) mit " o ". In Worten ausgedrückt, könnte man die Faktoren folgendermaßen beschreiben (es werden jeweils die Begriffe mit den höchsten Ladungen aufgeführt; in Klammern steht dahinter der mit dem Faktor erklärte Varianzanteil):

männliche Stichprobe:

FB1: klangvoll, klar, sympathisch (39.1%)

FB2: hell, dünn, hoch (15.8%)

FB3: ruhig, weich (11.7%)

FB4: gekünstelt, deutlich, langsam (8.1%)

weibliche Stichprobe:

FB1: bestimmt, aufgeweckt, frisch (40.2%)

FB2: weich, langsam, melodisch, ruhig (13.0%)

FB3: hoch, hell, dünn (11.5%)

FB4: gekünstelt, langsam, verständlich (7.6%)

Vertauscht man bei einem Geschlecht FB2 und FB3, so ergeben sich bei Betrachtung von Tabelle 4 sehr ähnliche Strukturen bei beiden Geschlechtern.

H A U P T K O M P O N E N T E N

Nr.	Attribut	männliche Stichprobe				weibliche Stichprobe			
		FB1	FB2	FB3	FB4	FB1	FB2	FB3	FB4
B1	ungezwungen	+	-	o	-	+	o	o	-
B2	sonor	+	-	o	o	+	o	-	o
B3	tief	o	-	o	o	o	+	-	o
B4	dünn	-	+	o	o	-	o	+	o
B5	gekünstelt	-	+	o	+	-	o	o	+
B6	angenehm	+	o	+	o	+	+	o	o
B7	kräftig	+	-	o	+	+	o	-	o
B8	deutlich	+	o	o	+	+	o	o	+
B9	langsam	o	+	+	+	-	+	o	+
B10	nüchtern	o	o	o	+	o	o	o	o
B11	hell	+	+	o	o	+	-	+	o
B12	aufgeweckt	+	o	-	o	+	o	o	o
B13	bestimmt	+	o	o	+	+	o	o	o
B14	ruhig	+	o	+	o	o	+	o	o
B15	sympathisch	+	o	+	o	+	+	+	-
B16	klar	+	+	o	o	+	o	o	o
B17	weich	o	+	+	o	o	-	o	o
B18	fest	+	o	o	+	+	-	o	o
B19	dumpf	-	-	o	o	-	+	-	o
B20	verständlich	+	+	o	o	+	o	o	+
B21	natürlich	+	o	o	-	+	o	o	+
B22	zaghaft	-	+	+	o	-	o	+	o
B23	klangvoll	+	o	o	o	+	+	o	o
B24	melodisch	+	+	o	o	+	+	o	o
B25	frisch	+	o	-	o	+	o	+	o

Tabelle 4:

Ladungen der Stimmbeurteilungsfaktoren (Hauptkomponentenanalyse):

" + ": stark positiv; " - ": stark negativ; "o": nur schwach (positiv oder negativ)

4.3.2.2. Darstellung von Einzelkorrelationen zwischen Stimmbeurteilungsfaktoren und akustischen Parametern

Möchte man Auskunft über akustische Korrelate auditiver Stimmbeurteilungen, so kann man zunächst alle Korrelationen mit den einzelnen akustischen Parametern berechnen.

Zur Vermeidung unübersichtlicher und umfangreicher Zahlentabellen wurden die Werte graphisch dargestellt. Die Abbildungen 39 bis 42 zeigen am Beispiel der weiblichen Stichprobe die akustischen Korrelate der vier Stimmbeurteilungsfaktoren gemäß dem in Abbildung 43 dargestellten Schema.

Es gibt zwar auf jeder Graphik nur wenige signifikante Werte, jedoch noch deutlich mehr, als durch reinen Zufall zu erwarten sind (siehe zum Vergleich Abbildung 34, Seite 92). Daß sich die meisten der signifikanten Werte auch interpretieren lassen sei im folgenden gezeigt:

Faktor FB1 (Abbildung 39):

Der Faktor FB1 kann als allgemein positiver Eigenschaftskomplex angesehen werden, mit Betonung auf "bestimmt", "aufgeweckt", "frisch".

Signifikante Korrelationen ergeben sich in einigen Werten von GSPK, F (außer F2), FAC9 und FAC19. In LSPK und seinen Derivaten ist kaum ein signifikanter Wert zu beobachten. (Zur Bedeutung der akustischen Parameter siehe Abbildung 26, Seite 70.)

FAC9 ist die zweite Hauptkomponente des Grundfrequenzspektrums. Daß aber nicht die Grundfrequenz selbst, sondern offensichtlich ein Muster im Spektralverlauf mit FB1 korreliert, sieht man an dem niedrigen Wert von F2 (mittlere Grundfrequenz). Was dieses Muster besagt, sieht man an GSPK und F3 bis F6:

F3 bis F6 drücken die Dynamik der Grundfrequenzschankungen aus (FAC19 ist die erste Hauptkomponente aus F3 bis F6). Daß diese Werte hohe positive Korrelationen mit FB1 aufweisen heißt, daß FB1 einen hohen Wert hat, wenn die Grundfrequenz stark beim Sprechen variiert.

Danach ist zu erwarten, daß im Grundfrequenzspektrum solcher Sprecher der Grundfrequenzgipfel eher flach und breit ist und nicht sehr stark ausgeprägt (da er in der Langzeitstruktur durch die starke Variation der Grundfrequenz verwischt wird). Genau dies zeigt sich bezüglich GSPK in Abbildung 39: Im Bereich der häufigsten Grundfrequenz (ca. 200 Hz bei Frauen) und ihrer ersten Harmonischen sind die Korrelationen negativ, dazwischen positiv.

Die negative Korrelation mit F1 (Lesezeit) sagt, daß sich in FB1 die Sprechgeschwindigkeit ausdrückt.

Kurz zusammengefaßt kann man sagen: Eine Sprecherin, die einen hohen Wert in Faktor FB1 hat, variiert die Grundfrequenz stark beim Sprechen und spricht relativ schnell.

Faktor FB2 (Abbildung 40):

Dieser Faktor beinhaltet in erster Linie die Attribute "weich", "langsam", "melodisch", "ruhig".

Die Bedeutung stark korrelierender akustischer Parameter sagt Folgendes:

LSPK und LSPKG: Die negative Korrelation mit Spektralwerten um 2 kHz erinnert an das Ergebnis der Voruntersuchung zur Sprechtonhöhe und Lautstärke (siehe die Zusammenfassung auf Seite 46): Die mittlere Intensität in diesem Bereich korreliert mit der Lautstärke.

F1: Positive Korrelation mit der Redezeit

F2: Negative Korrelation mit der Grundfrequenz (dies kommt auch in FAC8, der ersten Hauptkomponente des Grundfrequenzspektrums, zum Ausdruck sowie in GSPK selbst, indem dort in dem Bereich, in dem bei tiefen Frauenstimmen die Grundfrequenz liegt (und im Bereich der ersten Harmonischen davon), stark positive Werte zu finden sind.

RSPKI: Die hier beobachtbaren Korrelationen lassen sich nicht so leicht interpretieren; nachdem aber in RSPKIZ keine Signifikanz vorhanden ist, könnten sie mit der Sprechgeschwindigkeit in Zusammenhang stehen.

RSPKFZ: Positive Korrelationen liegen bei den höheren Frequenzen (ca. 2 Hz), negative bei den tieferen (ca. 0.5 Hz). Das heißt, daß bei hohen Werten von FB2 die Grundfrequenzschwankungen eher innerhalb von Wörtern als in der Satzintonation ausgeprägt sind.

Ein hoher Wert in FB2 beschreibt also eine leise, tiefe Stimme und eine langsame Sprechweise, bei der Grundfrequenzschwankungen mehr im Bereich um 2 Hz liegen als im Bereich um 0.5 Hz.

Faktor FB3 (Abbildung 41):

Der Faktor FB3 beinhaltet die Attribute "hell", "dünn", "nicht kräftig", "nicht tief", "nicht sonor".

Akustisch drückt sich dieser Faktor fast nur in Parametern aus, die mit der Grundfrequenz in Zusammenhang stehen (GSPK, F2, FAC8). Die signifikanten negativen Korrelationen in RSPKI liegen im Bereich schnell gesprochener Silben (ca. 5 Hz), lassen sich aber nicht sehr anschaulich interpretieren.

Der Faktor FB3 beinhaltet also im wesentlichen die Tonhöhe.

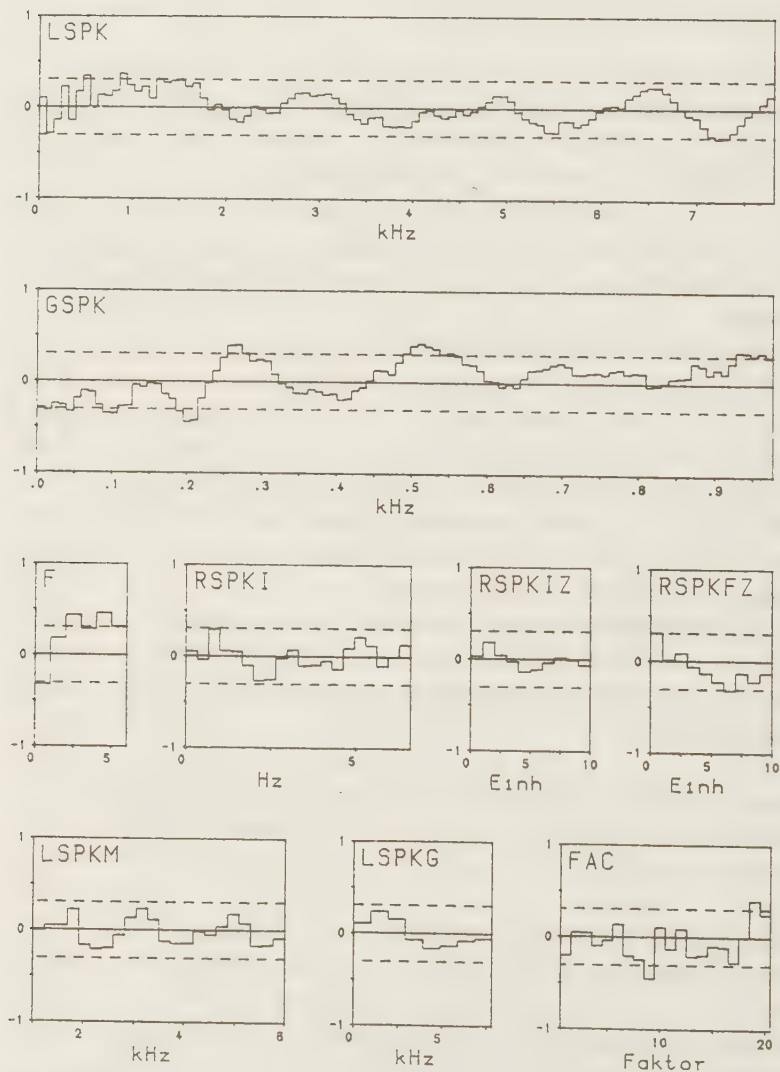


Abbildung 39:

Akustische Korrelate des Stimmbeurteilungsfaktors FB1 (weibliche Stichprobe):

Der Faktor bedeutet vorwiegend: "bestimmt", "aufgeweckt", "frisch"

(gestrichelt: Signifikanzgrenze für $p < 0.01$).

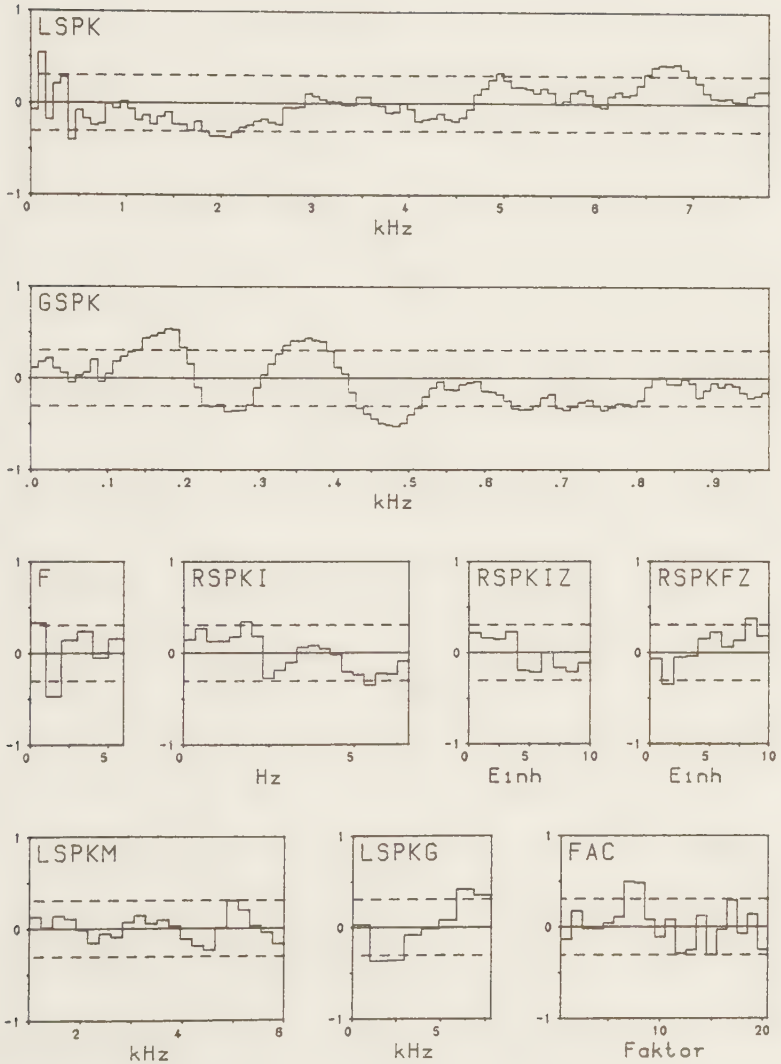


Abbildung 40:

Akustische Korrelate des Stimmbeurteilungsfaktors FB2 (weibliche Stichprobe):

Der Faktor bedeutet vorwiegend: "weich", "langsam", "melodisch", "ruhig"

(gestrichelt: Signifikanzgrenze für $p < 0.01$).

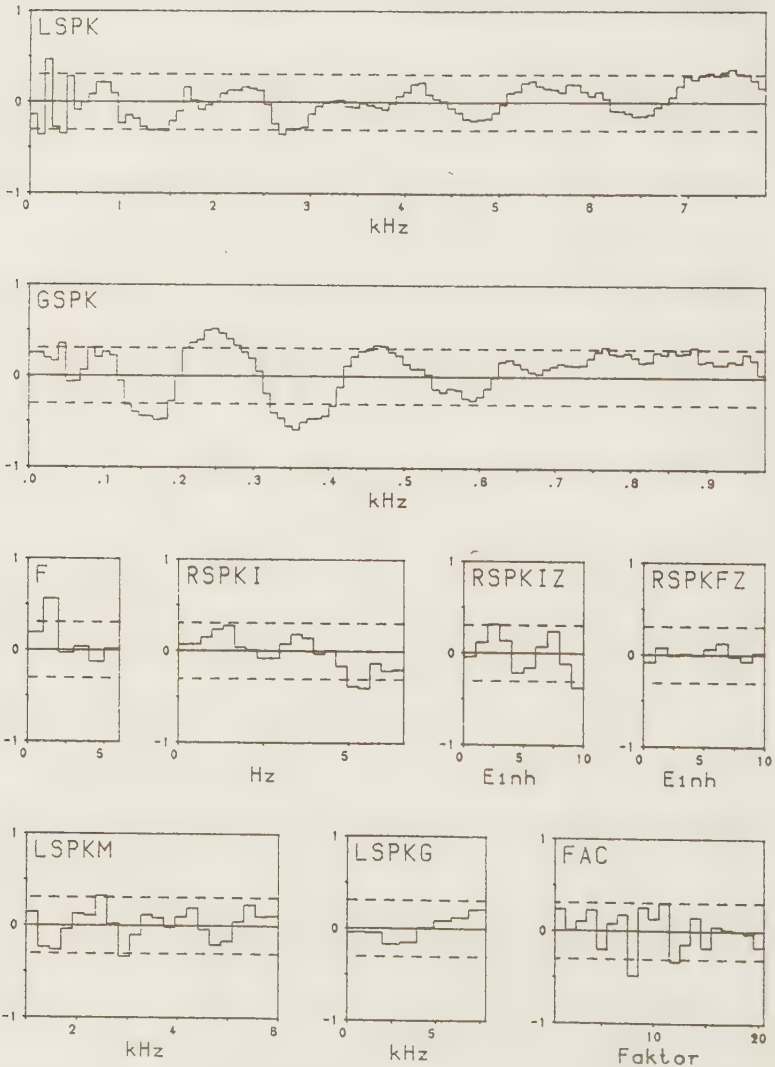


Abbildung 41:

Akustische Korrelate des Stimmbeurteilungsfaktors FB3 (weibliche Stichprobe):

Der Faktor bedeutet vorwiegend: "hoch", "hell", "dünn"

(gestrichelt: Signifikanzgrenze für $p < 0.01$).

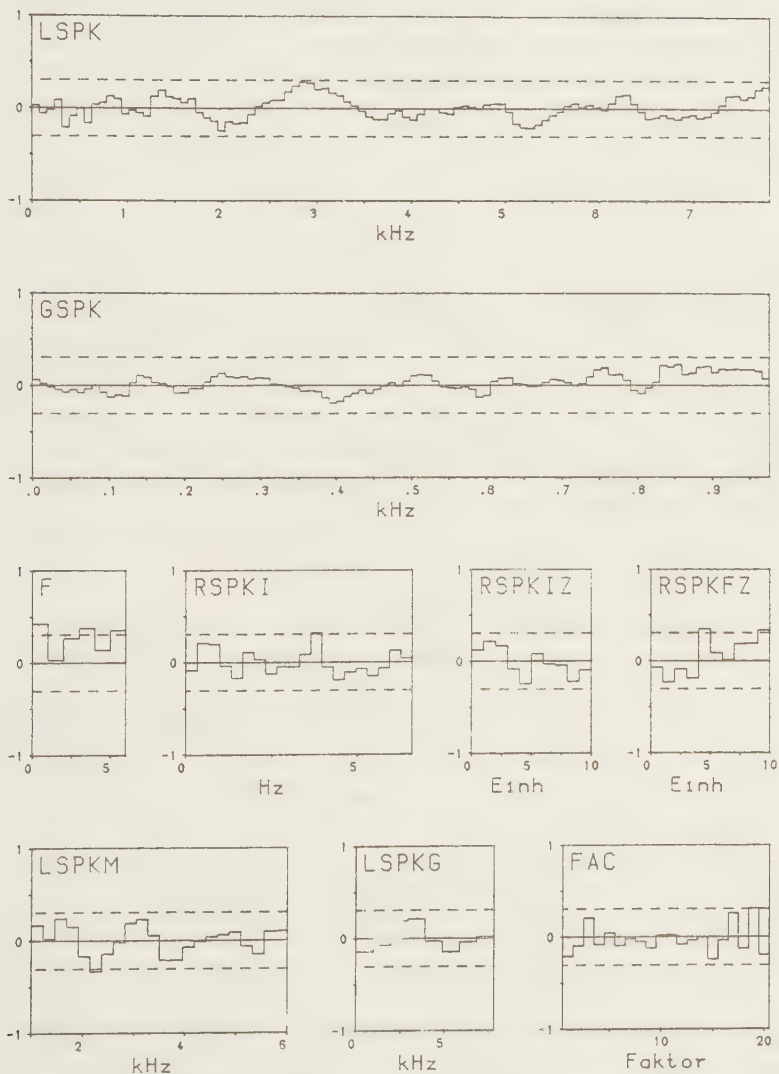


Abbildung 42:

Akustische Korrelate des Stimmbeurteilungsfaktors FB4 (weibliche Stichprobe):

Der Faktor bedeutet vorwiegend: "gekünstelt", "langsam", "verständnishaft"

(gestrichelt: Signifikanzgrenze für $p < 0.01$).

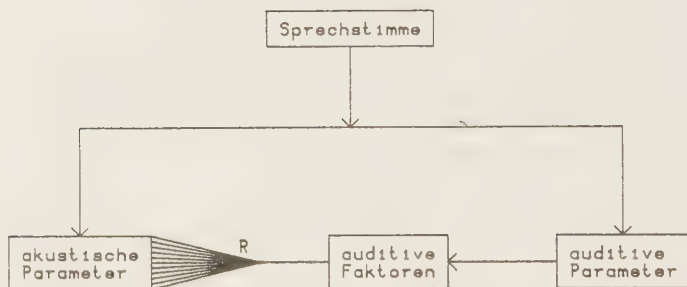


Abbildung 43:

Schema zur Berechnung der akustischen Korrelate auditiver Faktoren:

R: Menge der Einzelkorrelationen aller akustischen Parameter mit einem auditiven Faktor, wie sie in Abbildung 39 bis 42 dargestellt sind.

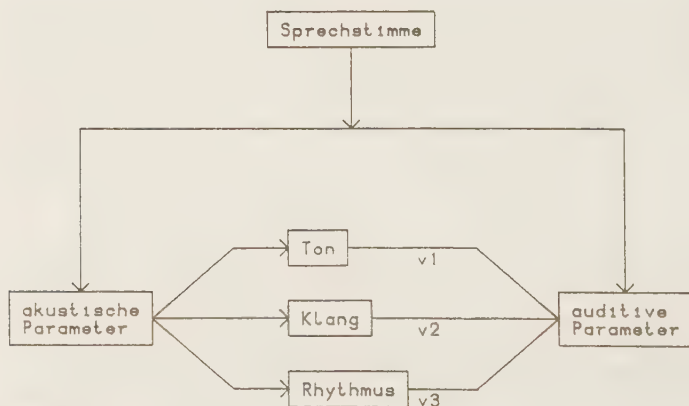


Abbildung 44:

Schema zur Ermittlung einer quantitativen Beziehung zwischen auditiver Stimmbeurteilung und drei akustischen Parametersätzen:

v1, *v2*, *v3*: Varianzanteil, der mit einer multiplen Regressionsfunktion, in die akustische Parameter eingehen, beschrieben werden kann (die konkreten Werte der einzelnen auditiven Attribute und Faktoren sind in Tabelle 5 aufgeführt).

Faktor FB4 (Abbildung 42): Der Faktor FB4 ("gekünstelt", "langsam", "verständlich") drückt sich akustisch hauptsächlich in einer niedrigen Sprechgeschwindigkeit (F1) und in der Betonung schnellerer Grundfrequenzschwankungen (mit einer Periodendauer von weniger als 1 Hz (F4, F6 und RSPKFZ) aus.

4.3.2.3. Abschätzung eines quantitativen Zusammenhangs zwischen auditiven und akustischen Parametern mit Hilfe multipler Regressionsfunktionen

Man kann sich die Frage stellen: Wie zuverlässig lassen sich die auf dem Stimmbewertungsfragebogen aufgeführten Begriffe aus objektiven akustischen Parametern schätzen?

Hierfür wurde die multiple Regressionsrechnung eingesetzt. Da aber die Stichproben nicht unendlich groß sind (männlich $n=47$, weiblich $n=70$), muß man sich für die Aufstellung multipler Regressionsgleichungen auf wenige Prädiktoren beschränken. Damit die Signifikanz der Regressionsfunktionen abgeschätzt werden kann, darf keine systematische Vorauswahl der Parameter getroffen werden; die eingehenden Werte müssen unabhängig von ihrer Korrelation mit dem zu schätzenden Merkmal bestimmt werden.

Einen kompakten Datensatz, der immerhin ungefähr 80% der gesamten Varianz in den akustischen Parametern beinhaltet, stellen die extrahierten Hauptkomponenten dar (FAC1 bis FAC20).

Eine Unterteilung in Untergruppen erscheint aus folgenden zwei Gründen angebracht:

- (1) Bei den gegebenen Stichprobengrößen ist es wenig sinnvoll, mehr als 10 Variablen gleichzeitig in eine Funktion eingehen zu lassen.
- (2) Es ist vielleicht von Interesse, nicht nur zu wissen, daß ein signifikanter Zusammenhang zwischen akustischen und auditiven Parametern besteht, sondern auch, in welchen Aspekten des Sprachsignals er lokalisiert ist.

Daher wurden aus dem akustischen Datensatz drei Prädiktorguppen zur Schätzung auditiver Attribute gebildet:

- (1) Klang: FAC1 bis FAC5 (Hauptkomponenten aus LSPKM, der mittleren Struktur des Langzeitspektrums oberhalb 1000 Hz), FAC6 und FAC7 (Hauptkomponenten aus LSPKG, der Grobstruktur des Langzeitspektrums);
- (2) Ton: F2 (mittlere Grundfrequenz), FAC8 bis FAC11 (Hauptkomponenten aus GSPK, dem Grundfrequenzspektrum bis 500 Hz), FAC17 und FAC18 (Hauptkomponenten aus RSPKFZ, dem Rhythmuspektrum des zeitkomprimierten Grundfrequenzverlaufs), FAC19 und FAC20 (Hauptkomponenten aus F3 bis F6, den Grundfrequenz – Streuungsmaßen);

(3) Rhythmus: F1 (Lesezeit), FAC12 bis FAC14 (Hauptkomponenten aus RSPKI, dem Intensitäts – Rhythmusspektrum), FAC15 und FAC16 (Hauptkomponenten aus RSPKIZ, dem Rhythmusspektrum des zeitkomprimierten Intensitätsverlaufs)

Die Untersuchung wurde nach dem in Abbildung 44 dargestellten Schema durchgeführt; v1, v2 und v3 sind die durch die Prädiktoren schätzbaren Varianzanteile der einzelnen Attribute bezüglich Klang, Ton und Rhythmus.

Die konkreten Werte sind in Tabelle 5 aufgeführt; die Sterne geben die Signifikanz der Regressionsfunktionen an (*: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$).

Um Tabelle 5 vernünftig interpretieren zu können, muß man sie auch mit Spalte 1 von Tabelle 3 (Seite 99) vergleichen, die Auskunft über die Zuverlässigkeit der auditiven Attribute selbst gibt, denn bei Attributen, bezüglich derer sich die Beurteiler nicht in ihren Wertungen einig sind, ist von vornherein kaum ein aussagekräftiges akustisches Korrelat zu erwarten; diese Begriffe sagen möglicherweise mehr über den Zuhörer als über den beurteilten Sprecher aus.

Einige Schlüsse, die sich aus den beiden Tabellen ziehen lassen, seien hier kurz zusammengefaßt:

(1) Daß bei der männlichen Stichprobe weniger signifikante Werte auftauchen als bei der weiblichen, liegt vermutlich nicht daran, daß es bei Männern weniger akustische Korrelate gibt, sondern daran, daß die Stichprobe kleiner und daher eine Signifikanz schwerer nachweisbar ist; die Zahlenwerte tendieren fast bei allen Merkmalen bei beiden Geschlechtern in dieselbe Richtung. Im folgenden sei daher nur die weibliche Stichprobe betrachtet.

(2) Folgende Attribute lassen sich aus den akustischen Parametern nicht oder nur schlecht schätzen: "angenehm", "deutlich", "nüchtern", "ruhig", "sympathisch", "verständlich", "natürlich", "melodisch".

Mit Ausnahme von "deutlich" und "natürlich" waren bei diesen Attributen auch die Korrelationen unter den Beurteilern niedrig.

(3) Alle Attribute, bei denen unter den Zuhörern gute Übereinstimmungen ($p < .001$) zu beobachten waren ("tief", "langsam", "hell", "aufgeweckt", "zaghaft", "frisch"), lassen sich auch objektiv gut schätzen.

(4) Die meisten Attribute lassen sich mit den Grundfrequenzparametern (v2) am besten

schätzen; das heißt, daß der Sprechtonhöhe und ihrem Verlauf für die Zuhörer offensichtlich die größte Bedeutung zukommt bei der Stimmbewertung.

(5) Das Langzeitspektrum oberhalb 1000 Hz stellt nur für die Attribute "kräftig" und "weich" einen zuverlässigen Satz von Prädiktoren dar. Interessant ist, daß die Begriffe "hell" und "dumpf" offensichtlich nicht mit dem Klang, sondern mit den Grundfrequenzparametern verknüpft sind.

Daß sich mehr sprachbezogene Attribute wie "deutlich" und "verständlich" nicht im Langzeitspektrum ausdrücken, kann als ein weiterer Hinweis angesehen werden, daß das Langzeitspektrum eher Sprecher- als sprachbezogene Eigenschaften erfaßt.

(6) Bezüglich der meisten Merkmale scheint die Schätzung durch den Computer durchaus dem Urteil eines Zuhörers äquivalent zu sein.

Attribut	männlich			weiblich		
	v1	v2	v3	v1	v2	v3
ungezwungen	.05	.26	.07	-.03	.19*	.09
sonor	-.02	.39**	-.04	.07	.27**	.07
tief	.05	.50***	.06	.10	.42***	.02
dünn	.02	.14	.04	.16	.29**	.00
gekünstelt	.04	.14	.14	-.02	.09	.28**
angenehm	-.08	.20	-.02	.03	-.02	.01
kräftig	.07	.20	.11	.26**	.32**	.05
deutlich	.03	.09	.04	.00	.06	-.01
langsam	.03	.21	.66***	.04	.39***	.48***
nüchtern	.12	.04	.15	-.08	-.04	.00
hell	.03	.37*	.00	.04	.39***	.01
aufgeweckt	.09	.24	.21	.01	.33**	.04
bestimmt	.19	.16	.10	.09	.27**	.10
ruhig	-.08	.10	.06	.07	.09	.09
sympathisch	-.06	.14	-.07	.05	-.11	.03
klar	-.03	.15	-.05	-.04	.29**	-.02
weich	.25*	.04	.21	.20*	.05	.03
fest	.09	.35*	.06	.12	.31**	.05
dumpf	.00	.18	.03	.08	.31**	.01
verständlich	-.04	.06	-.07	-.02	.10	-.02
natürlich	-.12	.19	-.01	.00	-.03	.12
zaghaf	.15	.39**	.25*	.09	.29**	.16*
klangvoll	-.08	.24	-.07	.09	.21*	-.07
melodisch	.00	.07	.00	.06	.16	-.06
frisch	.10	.11	.13	-.01	.31**	.07
FB1	-.03	.30*	.01	.03	.25*	.07
FB2	.18	.27	.18	.21*	.28**	.07
FB3	.07	.20	.26*	.10	.37***	.07
FB4	.15	.07	.23	.00	.12	.16*

Tabelle 5:

Aus objektiven Parametern schätzbarer Varianzanteil auditiver Attribute:

v1: Klangparameter; v2: Tonparameter; v3: Rhythmusparameter; (siehe Abbildung 44)

Signifikanzschwellen: *: $p < 0.01$; **: $p < 0.001$; ***: $p < 0.0001$

4.4. Sprechstimme und sonstige Sprechereigenschaften

4.4.1 Empirisch testbares Modell der nonverbalen Kommunikation

Die Tatsache, daß verschiedene Sprecher unterschiedlich sprechen, ermöglicht dem Zuhörer, einen Sprecher an der Stimme zu erkennen. Man kann sich nun fragen: Was erfährt der Zuhörer auditiv über einen Sprecher, wenn er ihn nicht kennt? Welches Bild macht er sich von ihm? Wie gut sind tatsächliche Eigenschaften des Sprechers in diesem Bild repräsentiert? Welche akustischen Parameter geben sie dem Zuhörer zu erkennen? Inwieweit kann auch der Computer mit Hilfe akustischer Parameter aus dem Sprachsignal Sprechereigenschaften erkennen?

Zur Beantwortung dieser Fragen wurde das in Abbildung 46 dargestellte Modell konstruiert, das eine gewisse Ähnlichkeit mit der von Scherer (1978) angewandten Version des Brunswick'schen Linsenmodells (Abbildung 45) aufweist. Im Gegensatz zu dem Modell von Scherer arbeitet das hier entwickelte Modell nicht mit akustischen und auditiven Einzelparametern sondern mit ganzen Variablenblöcken, aus denen multiple Regressionsfunktionen abgeleitet werden.

Die direkten Beziehungen zwischen akustischen und auditiven Parametern (bei Scherer "distal cues" und "percepts") sind hier ausgeklammert, da sie in Kapitel 4.3.2. gesondert betrachtet wurden.

Da das Modell auf verschiedene Eigenschaften angewendet wird, sind die Korrelationskoeffizienten nur allgemein angegeben (r_1 bis r_9); die Zahlenwerte finden sich in Tabelle 6.

Die Berechnung der Werte erfolgte in folgenden Schritten:

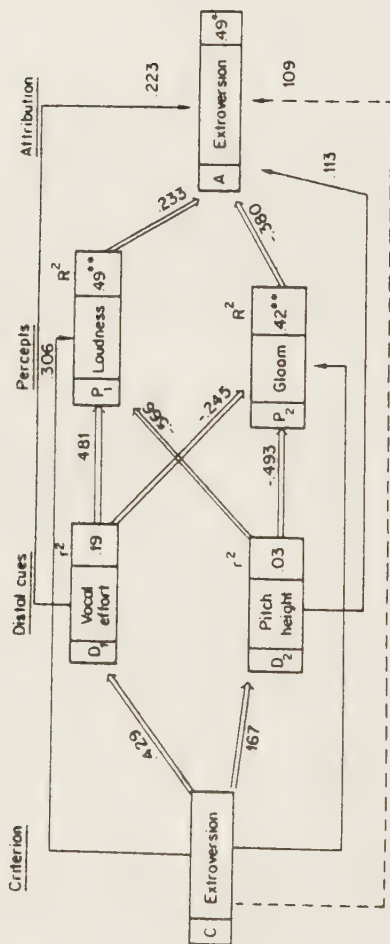
(1) Umformung sämtlicher Eigenschaften und Attributionen in Z-Werte (Mittelwert = 0, Standardabweichung = 1);

(2) Aufteilung der Versuchspersonen auf drei Gruppen (nach dem Zufall);

(3) Durchführung der Regressionsschätzungen:

Es wurden immer zwei der drei Gruppen zur Aufstellung von Regressionsgleichungen verwendet; für die verbleibende Gruppe wurden mit Hilfe dieser Gleichungen dann die Schätzwerte berechnet.

Dies wurde in allen drei Gruppenkombinationen durchgeführt, so daß für alle Sprecher empirisch überprüfbare Schätzwerte vorlagen.



Two-dimensional path analysis model for the extroversion inference

* $p < .05$ ** $p < .01$

† The dashed line represents the direct path, the double line the postulated indirect path, and the single lines the indirect paths compatible with the model. Coefficients shown are standardized coefficients except for r_{cp} and r_{cp} , which are Pearson r s (the direction shown is theoretically postulated). R^2 s based on all predictors from which paths lead to the variable.

Abbildung 45: (nach Scherer 1978)

Modell zur Darstellung der nonverbalen Kommunikation im Sprachsignal in Bezug auf das Persönlichkeitsmerkmal "Extrovertiertheit": Über Korrelationsanalysen wird hier versucht, "Pfade" zu finden, auf denen das Merkmal übermittelt wird.

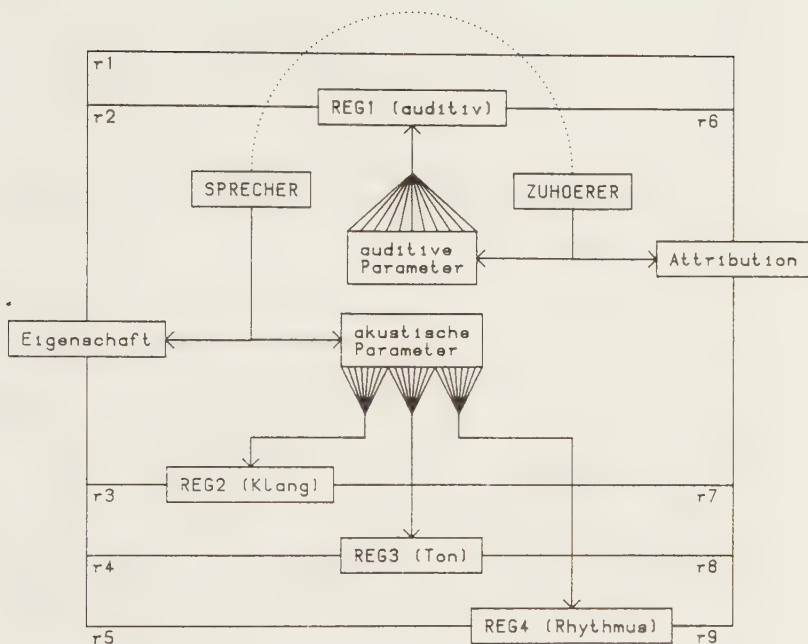


Abbildung 46:

Modell zu Darstellung von Zusammenhängen zwischen Sprechereigenschaften, Zuhörerattributionen und akustischen und auditiven Parametern aus dem Sprachsignal:

Aus den Parametern werden vier Regressionsfunktionen aufgestellt, mit denen die Sprechereigenschaft geschätzt wird. Ein Vergleich der tatsächlichen Eigenschaft mit der Attribution und der Regressions-Schätzwerte mit Eigenschaft und mit Attribution läßt Rückschlüsse darüber zu, inwieweit sich eine Sprechereigenschaft im Sprachsignal ausdrückt und wie sie übermittelt wird. Hierfür wurden für jede Eigenschaft 9 Korrelationskoeffizienten berechnet. Eine unmittelbare Prüfung auf Signifikanz ist möglich, da nicht dieselben Individuen zur Aufstellung der Regressionsfunktion und zur Schätzung verwendet wurden.

Die konkreten Werte für r_1 bis r_9 für die einzelnen Eigenschaften finden sich in Tabelle 6.

Die Schätzungen wurden mit folgenden Parametersätzen durchgeführt (siehe auch Abbildung 26, Seite 70, Abbildung 44, Seite 108 und Text Seite 109/110):

REG1: FB1 bis FB4 (Hauptkomponenten der auditiven Parameter)

REG2: FAC1 bis FAC7 (Hauptkomponenten aus dem Langzeitspektrum)

REG3: F2, FAC8 bis FAC11, FAC17 bis FAC20 (Grundfrequenz und ihr Verlauf)

REG4: F1, FAC12 bis FAC16 (Rhythmusparameter)

(4) Berechnung der 9 Korrelationskoeffizienten gemäß Abbildung 46 und der Signifikanz für den einseitigen Test (negative Werte werden nicht erwartet und sind als Unkorreliertheit zu betrachten).

Die Durchführung geschah mit den Prozeduren "CONDESCRIPTIVE", "REGRESSION" und "PEARSON CORR" des Programmpakets SPSSX (Schubö und Üehlinger 1984).

Auf folgende Merkmale wurde das Modell angewendet (hinter dem Schrägstrich steht jeweils das Attribut der auditiven Schätzung; siehe Anhang Seite A5):

- Lebensalter / geschätztes Alter
- neurotische Tendenz (PS2) / B40 – B41 – B42 + B43

(Dieser Begriff wurde, da er im allgemeinen Sprachgebrauch nicht sehr geläufig ist, durch die Zuhörerattribute "unsicher", "zufrieden", "ausgeglichen" und "gestört" angenähert.)

- Extrovertiertheit (PS1) / B38
- Körpergröße (K10) / B26
- Größenfaktor (K55) / FB5
- Längen – Breiten – Faktor (K56) / FB6

FB5 und FB6 sind zwei Faktoren, die aus einer Hauptkomponentenanalyse der auf den Körperbau bezogenen Schätzwerte (B26 bis B37) gewonnen wurden. Sie sind unmittelbar mit K55 und K56 vergleichbar:

In FB5 gehen alle Großheitsattribute (z. B. "groß", "breitschultrig", ...) positiv und alle Kleinheitsattribute (z. B. "schmächtig") negativ ein; in FB6 dagegen gehen die Langheits- (z. B. "langbeinig") und Schmalheitsattribute (z. B. "schlank") positiv und die Breitheitsattribute (z.B. "breitschultrig") negativ ein.

In Analogie zu K55 und K56 stellen also auch FB5 und FB6 einen Größen – und einen Längen/Breiten – Faktor dar.

Die (in Tabelle 6 aufgeführten) resultierenden Zahlenwerte für die in Abbildung 46 erklärten Korrelationskoeffizienten lassen sich folgendermaßen interpretieren:

Lebensalter:

r_1 ist hochsignifikant; das heißt, daß das Lebensalter auditiv gut geschätzt werden kann.

Daß r_2 niedriger ist als r_1 zeigt, daß sich in den auditiven Stimmbewertungsparametern das Lebensalter weniger ausdrückt als in der expliziten Altersschätzung.

Der hohe Wert von r_6 weist aber darauf hin, daß für die explizite Altersschätzung zum Teil Aspekte verwendet werden, die im Stimmbewertungsfragebogen erfaßt werden.

Bezüglich objektiver Stimmparameter drückt sich das Alter nur schwach in den Tonhöhenparametern aus, nicht im Klangspektrum und im Intensitätsrhythmus. (Diese Aussagen beziehen sich entsprechend der Stichprobe nur auf das Alter zwischen ca. 20 und 30 Jahren.)

Neurotische Tendenz und Extrovertiertheit:

Diese beiden mit dem "Eysenck Personality Inventory" erfaßten Persönlichkeitsmerkmale werden auditiv aus dem Sprachsignal kaum aufgenommen (unter den gegebenen Versuchsbedingungen). Weder die explizite Schätzung (r_1) noch die aus auditiven Stimmbewertungsparametern erstellte Regressionsfunktion (r_2) zeigt signifikante Werte.

In objektiven Parametern besteht lediglich bei der weiblichen Stichprobe bezüglich Extrovertiertheit eine leichte Korrelation im Klangspektrum (r_3) und im Sprechrhythmus (r_5). Bei den Männern drückt sich im Klangspektrum die neurotische Tendenz schwach aus (r_3). Daraus, daß r_8 einen relativ hohen Wert hat, kann man schließen, daß die Zuhörer nur den noch schwächeren Zusammenhang mit den Tonhöhenparametern nutzen, um das Merkmal zu schätzen, nicht aber das Klangspektrum.

Körpergröße:

Die hohen Werte für r_1 und r_3 zeigen, daß sich die Körpergröße sowohl für den Zuhörer als auch in objektiven Parametern (im Klangspektrum) im Sprachsignal ausdrückt.

Aus der Tatsache, daß r_7 sehr hoch ist, kann man schließen, daß die Schätzungen von Computer und Zuhörer oft parallel gehen (unabhängig von der Richtigkeit der Schätzungen), das heißt, daß der Zuhörer implizit die Regressionsbeziehung bezüglich des Klangspektrums benutzt.

In den Tonhöhenparametern und im Sprechrhythmus drückt sich die Körpergröße nicht aus.

Größenfaktor:

Ein ähnliches Muster wie bei der Körpergröße findet man bei der ersten Hauptkomponente der anthropometrischen Daten. Dies ist insofern verständlich, als diese beiden Maße sehr stark miteinander korrelieren.

Längen – Breiten – Faktor:

Die zweite Hauptkomponente der anthropometrischen Merkmale, in die im wesentlichen Längenmaße positiv und Breitenmaße negativ eingehen (die also den größenunabhängigen Formfaktor mit dem größten Varianzanteil darstellt), zeigt keine signifikante Korrelation mit der Sprechstimme (weder in auditiven noch in objektiven akustischen Parametern).

Das Modell wurde noch auf weitere Merkmale angewendet (z. B. Unterarmumfang, Extremitätenlängen, anthropometrische Indices ...) Dabei zeigte sich, daß bei absoluten Größenmerkmalen (z. B. auch bei der Handgröße) immer mehr oder weniger hohe Korrelationen zu finden waren, bei reinen Formmerkmalen jedoch nicht.

Diese Beobachtung wird auch bestätigt durch das Ergebnis aus Kapitel 4.2.5. (Seite 89), das zeigt, daß sich die Andromorphie innerhalb eines Geschlechts bezüglich anthropometrischer Formmerkmale nicht in der Stimme abbildet.

	Lebensalter		neurotische Tendenz		Extrovertiert- heit	
	männl.	weibl.	männl.	weibl.	männl.	weibl.
r1	.63***	.54***	.24	.00	.01	.00
r2	.17	.35**	.00	.02	.05	.00
r3	.01	.00	.30*	.00	.00	.28*
r4	.16	.25*	.22	.02	.00	.00
r5	.16	.01	.00	.00	.24	.30**
r6	.27*	.44***	.00	.00	.00	.03
r7	.03	.05	.12	.18	.00	.02
r8	.18	.24*	.27*	.00	.00	.13
r9	.03	.00	.00	.08	.06	.00

	Körpergröße		1. anthrop. Hauptkomponente		2. anthrop. Hauptkomponente	
	männl.	weibl.	männl.	weibl.	männl.	weibl.
r1	.28*	.46***	.50***	.31**	.22	.00
r2	.07	.00	.36**	.00	.00	.00
r3	.41**	.40***	.33*	.45***	.11	.00
r4	.00	.06	.14	.06	.00	.08
r5	.01	.00	.14	.00	.03	.00
r6	.33*	.32**	.74***	.50***	.00	.00
r7	.31*	.52***	.22	.44***	.07	.16
r8	.00	.00	.00	.06	.00	.17
r9	.00	.06	.23	.06	.16	.00

Tabelle 6:

Korrelationen zwischen Sprechstimme, auditiver Beurteilung und Sprechereigenschaften: Die neun Korrelationskoeffizienten wurden nach dem Modell in Abbildung 46 berechnet. An den Mustern der signifikanten Werte kann man erkennen, ob sich ein Merkmal in der Stimme ausdrückt und ob der Zuhörer für seine Schätzung Parameter benutzt, die auch objektiv zugänglich sind.

Negative Werte wurden gleich "0" gesetzt, da sie nach dem Modell nicht erwartet werden und nur die Zufallsstreuung der Stichprobe bei Unkorreliertheit der verglichenen Größen zeigen. (*: $p < .05$; **: $p < .01$ ***: $p < .001$)

4.4.2. Genauere Betrachtung der Korrelationen von Sprechereigenschaften mit einzelnen akustischen Parametern am Beispiel der Körpergröße

Möchte man nicht nur wissen, ob sich die Körpergröße im Klangspektrum ausdrückt, sondern auch, wie sie sich darin ausdrückt, so muß man sämtliche akustischen Parameter einzeln betrachten. Zu diesem Zweck wurden zwei Graphiken erstellt (Abbildung 47 und 48); die erste gibt Auskunft über die Korrelationen mit der Körpergröße (bei beiden Geschlechtern), die zweite (am Beispiel der weiblichen Stichprobe) darüber, wie sich die von den Zuhörern geschätzte Körpergröße (B26) im Gegensatz zur gemessenen abbildet.

Abbildung 47 zeigt, daß die Körpergröße in erster Linie mit dem Langzeitspektrum oberhalb 1000 Hz korreliert; das Muster ist bei Männern und Frauen (trotz des ausgeprägten Geschlechtsdimorphismus) ähnlich. (LSPK, LSPKM und LSPKG; aus FAC1 bis FAC7 kann diese Parallele nicht erkannt werden, da die Hauptkomponenten bei Männern und Frauen getrennt extrahiert wurden und daher unterschiedliche Ladungsmuster aufweisen.)

Aus Abbildung 48 geht hervor, daß die Zuhörer offensichtlich für ihre Schätzung genau dasselbe Muster benutzen, das sich auch objektiv im Klangspektrum ausdrückt (LSPK, LSPKM und LSPKG). Eine Diskrepanz dagegen zeigt sich in der Sprechtonhöhe (F2 und GSPK): sie korreliert mit der Körpergröße überhaupt nicht, der Zuhörer nimmt aber eine Negativkorrelation zwischen Tonhöhe und Körpergröße an. (Bei Männern ist dieser Effekt noch ausgeprägter.)

Eventuell spielt dafür die ganz allgemeine Vorerfahrung eine Rolle, daß z. B. große Tiere, Autos oder Maschinen im Schnitt tiefere Töne und Geräusche produzieren als kleine.

Im Sprechrhythmus finden sich keine hoch signifikanten Korrelationen mit der Körpergröße; eine Parallelität der Linien in Abbildung 48 (RSPK1, RSPKIZ und RSPKFZ) ist daher auch nicht zu erwarten.

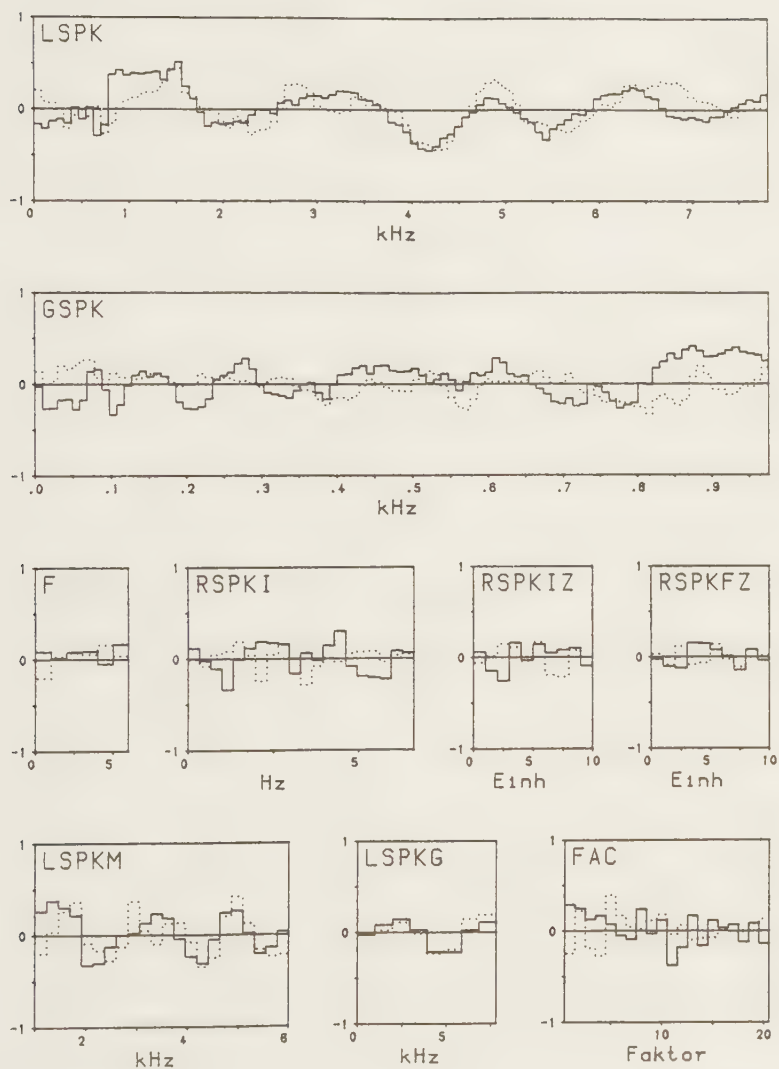


Abbildung 47:

Korrelationen zwischen Körpergröße und akustischen Einzelparametern:

durchgezogene Linie: männliche Stichprobe

gepunktete Linie: weibliche Stichprobe

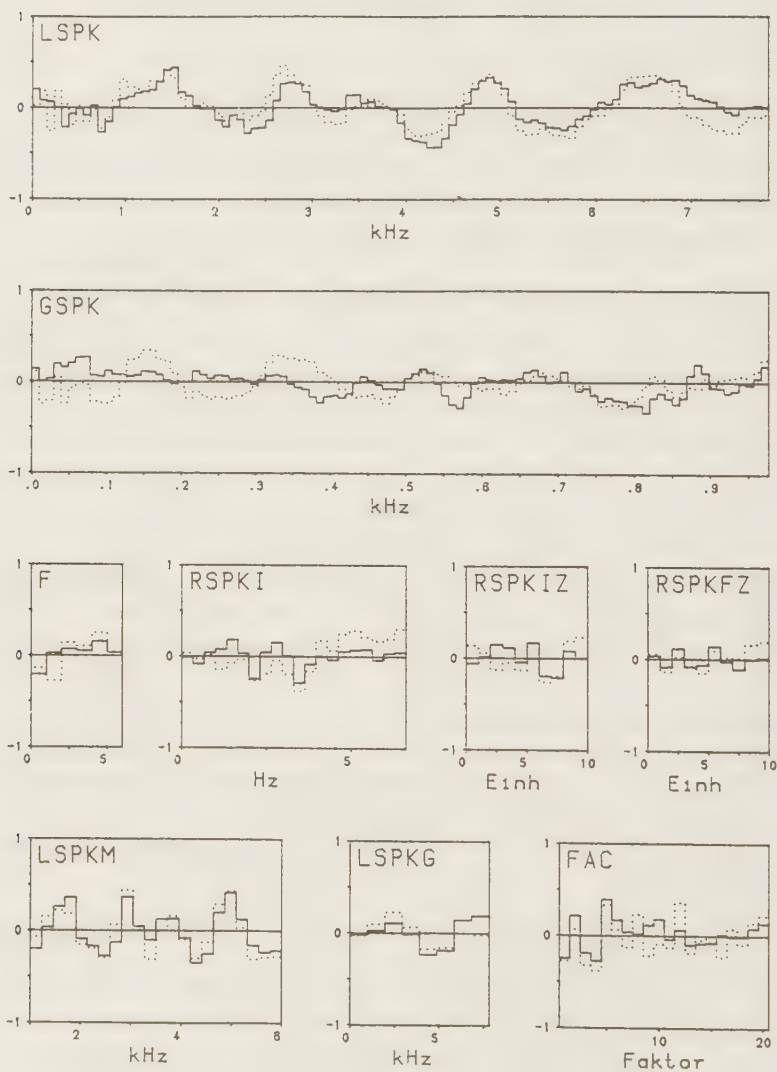


Abbildung 48:

Korrelationen akustischer Parameter mit der Körpergröße (durchgezogene Linie) und mit der Zuhörerattribut "groß" (gepunktet):

Vor allem im Langzeitspektrum oberhalb 1000 Hz (LSPK und LSPKM) zeigt sich eine gute Übereinstimmung.

5. Diskussion

In diesem Kapitel werden die wesentlichen Punkte der Arbeit, was die Methodik und die Ergebnisse anbelangt, diskutiert und, soweit möglich, mit Beispielen aus der Literatur verglichen.

5.1. Methodik

5.1.1. Digitale Berechnung der Langzeitspektren

Die gebräuchlichste Methode zur Berechnung von Langzeitspektren ist die Verwendung eines Terzfilter-, $1/2$ -Oktavfilter oder Oktavfilter-Analysators (z. B. Hargreaves und Starkweather 1963). Diese Analysatoren arbeiten logarithmisch (das heißt, die Frequenzbänder werden, in Hz gemessen, mit steigender Frequenz breiter); das Frequenz-Auflösungsvermögen ist vorgegeben und kann nicht beliebig variiert werden.

Das hier eingesetzte digitale Verfahren, mit dem das Langzeitspektrum durch schrittweise Fourier-Transformation kurzer Signalstücke und anschließender Mittelung über die Einzelspektren berechnet wird, hat den Vorteil, daß je nach Fragestellung die Frequenzauflösung variiert werden kann. So kann zum Beispiel eine Auflösung von 10 Hz gewählt werden, so daß im unteren Spektralbereich die mittlere Grundfrequenz gut als Gipfel erkannt werden kann (siehe Abbildung 19, Seite 56). Ein Terzfilter-Analysator dagegen hat im Bereich der Grundfrequenz nur ein Auflösungsvermögen zwischen 30 und 70 Hz.

Vor allem in höheren Frequenzbereichen, in denen das Auflösungsvermögen mit logarithmischer Skala ja niedriger wird, bietet das hier verwendete Verfahren Vorteile. Als Beispiel kann der in Abbildung 27 (Seite 76) dargestellte Geschlechtsdimorphismus betrachtet werden: In der mittleren Struktur des Langzeitspektrums (mit 240-Hz-Bändern) oberhalb 1000 Hz (LSPKM) zeigt sich zwischen 3000 und 4000 Hz ein ausgeprägter Dimorphismus. In diesem Bereich ist das Auflösungsvermögen eines Terzfilter-Analysators bereits gröber als 1000 Hz und würde den Gipfel bei ca. 3500 Hz nicht mehr scharf abbilden.

Ein weiterer Vorteil der digitalen Analyse ist die einfache Abspeicherung der Ergebnisdaten. Eine manuelle Eingabe von Zahlenwerten, bei der leicht Fehler passieren, ist überhaupt nicht mehr erforderlich, die Spektralwerte können zur Berechnung sekundärer Parameter oder für eine statistische Auswertung jederzeit maschinell aus der

Speicherdatei eingelesen werden.

5.1.2. Parameterextraktion aus dem Sprechrhythmus

Zur Gewinnung von Parametern aus dem Sprechrhythmus ist in der Literatur noch kein vollautomatisches Verfahren in diesem Umfang beschrieben. Die den Rhythmusparametern zugrundeliegenden Verlaufskurven (Intensitätsverlauf und Grundfrequenzverlauf) können zwar auch mit nicht-digitalen Verfahren gewonnen werden, eine sekundäre Verarbeitung dieser Kurven beschränkte sich aber bislang auf die Ermittlung von Größen wie Sprechgeschwindigkeit und Varianz von Intensität und Grundfrequenz.

Besonders der Umgang mit dem Grundfrequenzverlauf bereitete immer wieder Schwierigkeiten. Obwohl es einige Algorithmen gibt (z. B. Autokorrelations- oder Cepstralanalyse), ihn zu extrahieren (Heß 1983), gibt es keinen, der auf alle Sprecher in gleicher Weise anwendbar ist und der eine weitgehend stetige Kurve längerer Sprechproben aufzeichnet.

Hier wurde ein Verfahren entwickelt, bei dem einfach der Grundfrequenzgipfel aus Einzel-Fourierspektren ermittelt wird, was den Vorteil hat, daß die Grundfrequenz selbst mit hinreichend hoher Intensität vorhanden sein muß, um als solche akzeptiert zu werden. Mit einer zusätzlichen Entscheidungslogik werden unplausible Werte eliminiert und fehlende interpoliert. Nur bei einer von 120 Sprechproben arbeitete das Verfahren nicht einwandfrei, was vom System selbst aufgrund einer ungewöhnlich hohen Anzahl unplausibler Werte registriert wurde.

Eine nachfolgende Berechnung von Rhythmusspektren aus den Verlaufskurven wurde bislang, abgesehen von Gregory und Hoyt (1982), die auf Papier aufgezeichnete Intensitätskurven halbmanuell weiterverarbeiteten, um Gesprächspartneradaptationen im Sprechrhythmus nachzuweisen, noch nicht durchgeführt; es gibt also noch kein Vergleichsmaterial das Aufschluß über die Relevanz solcher Rhythmusspektren gibt.

Die hier beschriebenen Beobachtungen zum Geschlechtsdimorphismus im Sprechrhythmus und zu auditiven Korrelaten der Rhythmusparameter zeigen immerhin, daß die Rhythmusspektren nicht arbiträr zustandekommen, sondern daß mit ihnen Sprechereigenschaften erfaßt werden.

5.1.3. Datenreduktion durch Hauptkomponentenanalyse

Die Hauptkomponentenanalyse stellt ein bequemes Verfahren dar, einen redundanten Datensatz auf wenige aussagekräftige Werte zu reduzieren. Sie wird zum Beispiel in der Psychologie zur Verarbeitung von umfangreichen Fragebogendaten eingesetzt, aber auch in anderen Bereichen, in denen es darauf ankommt, viele Einzelparameter auf wenige relevante Größen zu reduzieren (z. B. auch in der Meteorologie). Oft wird der Faktorraum nachträglich "rotiert", damit die resultierenden Faktorwerte im Sinne unseres sprachlichen Begriffsrasters interpretierbar werden.

In der vorliegenden Arbeit wurde die Hauptkomponentenanalyse zur Gewinnung sekundärer Parameter aus dem Sprachsignal, zur Zusammenfassung auditiver Stimm- und Sprecherbeurteilungen und zur Verarbeitung anthropometrischer Meßdaten verwendet. Auf eine Rotation der Faktoren wurde aus folgenden Gründen verzichtet:

- (1) Bei den akustischen Spektralwerten gibt es keinen Anhaltspunkt für eine solche Rotation, sie wäre nur eine willkürliche Operation.
- (2) Bei den anthropometrischen Merkmalen sind bereits die Hauptkomponenten selbst anschaulich interpretierbar (Größen- und Längen/Breiten-Faktor). Die Relevanz dieser Hauptkomponenten wird erstens dadurch angezeigt, daß die Faktor- Ladungsmuster bei Männern und Frauen fast dieselben sind, zweitens dadurch, daß sich in den auf die Konstitution des Sprechers bezogenen auditiven Beurteilungsdaten eine unmittelbar vergleichbare analoge Struktur zeigt. (Das heißt, der Zuhörer hat in sich eine Vorstellung von der Korreliertheit und der Varianz von Körpermaßen, die mit der objektiv gemessenen und berechneten übereinstimmt.)
- (3) Auch bei den Stimmbeurteilungsparametern kann man die Hauptkomponenten interpretieren, auch dort ergeben sich bei Männern und Frauen weitgehend parallele Ladungsmuster (siehe Tabelle 4, Seite 101).

Als empirischen Hinweis für die Brauchbarkeit der Hauptkomponenten bei der Durchführung von Diskriminanz- und multiplen Regressionsanalysen können die Ergebnisse zur automatischen Geschlechtserkennung (Tabelle 2, Seite 86) und zur Schätzung von Sprechereigenschaften (Tabelle 6, Seite 119) angesehen werden.

5.1.4. Smallest Space Analysis

Das Verfahren der "smallest space analysis" (Prozedur SSA1 in SPSS, Toledano 1977) ist eine multidimensionale Skalierungsmethode, die für viele Fragestellungen eingesetzt werden kann. Sie stellt keine hohen Anforderungen an die zu bearbeitenden Daten (es genügt Ordinalskalen – Niveau) und ermöglicht, komplizierte Zusammenhänge schnell und anschaulich darzustellen. Als Eingabe braucht man lediglich eine symmetrische quadratische Distanz – oder Ähnlichkeitsmatrix von $n \times n$ Elementen (Merkmale oder Individuen). In der vorliegenden Arbeit wurde das Verfahren sowohl zur Darstellung von Individuenähnlichkeiten als auch von Merkmalskorrelationen eingesetzt. Für die graphische Darstellung der zweidimensionalen Verteilungsmuster wurde ein neues Graphikprogramm erstellt, das ermöglicht, die Elemente im Raum nachträglich zu markieren.

Am Beispiel des Geschlechtsdimorphismus sei kurz gezeigt, daß die smallest space analysis als Methode zur Gruppentrennung durchaus mit der Diskriminanzanalyse verglichen werden kann, ihr bei kleinem Stichprobenumfang sogar deutlich überlegen ist: Abbildung 28 (rechts) zeigt die Sprecherähnlichkeiten im Grundfrequenzspektrum; es lassen sich keine Überlappungen zwischen den Geschlechtern feststellen. Angenommen, das Geschlecht sei auf der Abbildung nicht markiert, so könnten wir dennoch an der breiten Kluft zwischen den Geschlechtern erkennen, daß sich die Gesamtheit in zwei Gruppen untergliedern läßt. Legen wir nun eine Gerade in diese Kluft, so brauchen wir in jeder der beiden Graphiken nur das Geschlecht eines einzigen Sprechers zu wissen, um eine vollständige Zuordnung zu erhalten. In der oberen Abbildung würden wir so vermutlich eines der beiden Individuen, die in der Kluft liegen, falsch einordnen, in der unteren Abbildung alle richtig; wir hätten also eine Erkennungsrate von ca. 99%, das heißt eine ebenso hohe wie sie uns die Diskriminanzanalyse im besten Fall liefert. Letztere jedoch benötigt zur Aufstellung der Diskriminanzfunktion nicht 2 sondern mindestens ca. 50 Individuen, um zu dieser Erkennungsrate zu gelangen.

Abbildung 22 zeigt, daß mit der smallest space analysis bereits bei einer Stichprobe von 20 Individuen eine absolute Geschlechtertrennung gezeigt werden kann; die Anwendung der Diskriminanzanalyse wäre hier aufgrund der wenigen Freiheitsgrade nicht möglich oder zumindest problematisch.

Die Tatsache, daß die Prozedur SSA1 bei großen Stichproben lange Rechenzeiten benötigt (bei 60 Individuen ca. 45 Sekunden auf der CDC – Rechenanlage Cyber 875), dürfte mit zunehmender Verfügbarkeit von Großrechnern kein Problem darstellen.

In der Anthropologie könnte das Verfahren gut für den multivariaten Vergleich kleiner Populationen eingesetzt werden.

5.2. Diskussion der Ergebnisse im Hinblick auf verschiedene Fragestellungen

5.2.1. Sprecheridentifizierung

Zur Sprecheridentifizierung können verschiedene Ergebnisse früherer Studien bestätigt werden:

Das Langzeitspektrum ist grundsätzlich für eine Sprecheridentifizierung geeignet; die Identifikationsraten dürften bei dem hier verwendeten Verfahren noch etwas höher liegen als die von Pruzansky (1963), Hargreaves und Starkweather (1963) und Clarke und Becker (1969) berichteten (je nach Versuchsanordnung und Sprecherstichprobe 80 – 100%).

Auch die Äquivalenz verschiedener Vergleichsmaße für Langzeitspektren kann bestätigt werden: Majewski und Hollien (1975) und Zalewski, Majewski und Hollien (1975) kamen zu gleich guten Ergebnissen mit Korrelationskoeffizient und Euklidischem Abstand. In dieser Studie wurde der Korrelationskoeffizient mit X^2 , einem dem Euklidischen Abstand ähnlichem Maß, verglichen; mit beiden Maßen ließen sich gleich gute Erkennungsraten erzielen.

Eine schlechte Identifizierbarkeit verstellter Stimmen berichten Hollien und Majewski (1977): Diese Stimmen konnten nur zu ca. 30% erkannt werden (gegenüber 92 – 100% bei nicht verstellten Stimmen).

Die auf Seite 53/54 aufgeführten Ergebnisse zur Identifikation des Stimmimitators mit Hilfe der Diskriminanzanalyse liegen ebenfalls in dieser Größenordnung. Offensichtlich ist also das Langzeitspektrum willentlich beeinflussbar (wenn auch nicht sehr leicht, wie die Untersuchung des Einflusses von Lautstärke und Tonhöhe zeigt, der allein noch nicht ausreicht, die Identifikationsrate entscheidend herabzusetzen). Die Tatsache, daß der Imitator aber nicht in die mit objektiven Methoden ermittelte intraindividuelle Streuungsbreite des Imitierten gelangt, zeigt allerdings, daß es nicht leicht ist, das Langzeitspektrum gezielt in eine bestimmte Richtung zu verändern; es ist also ein Merkmalssystem, das der willentlichen Kontrolle der Sprechers weitgehend unzugänglich ist. Dies wird auch durch die Studien von Luck (1969) und Doddington (1973) bestätigt, in denen die Imitatorakzeptanz zwar noch leicht über dem Zufall lag, aber dennoch sehr gering war, gemessen an der normalen Identifikationsrate.

5.2.2. Sprechlautstärke und Tonhöhe

Die Voruntersuchung zum Einfluß von Sprechlautstärke und Tonhöhe wurde hauptsächlich zu dem Zweck durchgeführt, Merkmale zu finden, die stabil sind gegen diese beiden Einflußgrößen.

Daß sich im Spektrum leichte Veränderungen ergeben können, zeigt Abbildung 9 (Seite 40). Die beobachtete Anhebung der Intensität im Bereich zwischen 2000 und 3000 Hz mit Zunahme der (vom Sprecher intendierten) Sprechlautstärke, findet eine Bestätigung bei Kiukaanniemi (1982). Auch Winner (1983) beobachtete spektrale Veränderungen in Abhängigkeit von Lautstärke und Tonhöhe; allerdings untersuchte sie Einzellaute; daher sind die Ergebnisse nicht unmittelbar vergleichbar.

Verglichen mit der interindividuellen Varianz im Langzeitspektrum ist allerdings der Einfluß von Tonhöhe und Lautstärke relativ gering. Insbesondere die mittlere Struktur des Langzeitspektrums ist davon wenig beeinflusst, wie man an Abbildung 12 und 13 (Seite 44) sehen kann.

5.2.3. Geschlechtsdimorphismus

Nicht nur in der Sprechtonhöhe, sondern auch in höheren Spektralbereichen zeigt sich ein ausgeprägter Geschlechtsdimorphismus. (Hinweise hierfür zeigte auch Jacobshagen 1983.) So kann allein aus den ersten 5 Hauptkomponenten der mittleren Struktur des Langzeitspektrums das Geschlecht unbekannter Sprecher in 94% aller Fälle richtig ermittelt werden. Diese Erkennungsrate ist nur geringfügig schlechter als die nach anthropometrischen Formmerkmalen mit Hilfe der Diskriminanzanalyse (siehe Tabelle 2, Seite 86).

Auch im Sprechrhythmus sind signifikante Geschlechterunterschiede nachweisbar.

Eine Parallele ergibt sich hier zu den Ergebnissen einiger auditiver Studien, in denen die Geschlechtererkennung aus Signalen, die die Sprechtonhöhe nicht enthalten, getestet wurde:

Spielt man Zuhörern nur die Formanten (Maxima im Frequenzspektrum oberhalb der Grundfrequenz, deren Muster zur Unterscheidung verschiedener stimmhafter Sprachlaute dient) von Vokalen vor, so ergibt sich eine Erkennungsrate von ca. 80% (Coleman 1971). Ähnliche Ergebnisse berichten Ingemann (1968) und Schwartz (1968), die als Signal stimmlose Frikative benutzten und Schwartz und Rine (1968) mit geflüstelter Sprache.

Nach Lass (1976) liegt die auditive Geschlechtererkennung normaler Sprache bei 96 – 99%,

(was hier bestätigt werden kann, siehe Tabelle 2, Seite 86), 255-Hz-Hochpaß-gefilterter Sprache bei 91% und geflüsterter bei 75%.

Gemessen an diesen Zahlen, scheinen die in der vorliegenden Arbeit angewandten Methoden zur automatischen Geschlechterkennung dem menschlichen Zuhörer leicht überlegen zu sein.

Zwei männliche Sprecher, deren Geschlecht auditiv nicht erkannt wurde, zeigten auch in objektiven akustischen Parametern weibliche Merkmale.

Die Tatsache, daß diese beiden Sprecher auch in anthropometrischen Merkmalen an der Grenze zur weiblichen Variabilität lagen, gab Anlaß zu der Vermutung, daß generell ein Zusammenhang zwischen Maskulinität in Stimme und Maskulinität im Körper (innerhalb einer Geschlechtergruppe) besteht. Aufgrund der Ergebnisse aus Kapitel 4.2.5. (Seite 89) muß dieser Gedanke jedoch verworfen werden; für die normale Variabilität von Stimme und Körper ist kein solcher Zusammenhang nachweisbar.

Vergleicht man nun die Geschlechterunterschiede in der menschlichen Sprechstimme mit denen in den Stimmen von Tieren, so scheint der extreme Dimorphismus in der Grundfrequenz ein typisch menschliches Merkmal zu sein. Bei allen übrigen Säugern ist er geringer (Damste 1967).

Als eine mögliche Erklärung hierfür könnte man Folgendes anführen:

Daß das Geschlecht sich überhaupt in der Stimme ausdrückt, ist durch die große Bedeutung akustischer Signale für das Sozialverhalten leicht erklärbar. Nun kann man sich fragen: Wie geben unsere Verwandten, die Affen, ihr Geschlecht akustisch zu erkennen und warum machen wir es anders?

Im Gegensatz zur menschlichen Sprache haben Affen ein sehr begrenztes Repertoire an Lautgebungen, das weitgehend genetisch programmiert ist (siehe zum Beispiel Herzog und Hopf 1984) und dessen Bedeutungsstruktur für die jeweilige Art (zum Teil sogar artübergreifend) festgelegt ist. So wurden für Colobus 6 unterschiedliche Grundlaute beschrieben, für Gibbons 12 – 15, für Schimpansen 13 und für Gorilla 16 (Marler 1978).

Das Geschlecht wird hauptsächlich dadurch signalisiert, daß Männchen und Weibchen unterschiedliche Laute produzieren. Eine genaue Aufzeichnung solcher geschlechtsspezifischer Laute bei Gibbons (*Hylobates concolor leucogenys*) findet sich bei Schilling (1984). Bei den Affen unterscheiden sich die Geschlechter also in dem, w a s sie sagen und weniger in dem, w i e sie es sagen.

Mit zunehmender Komplexität des Vokabulars und seiner Verknüpfungsmöglichkeiten und mit der Abkoppelung der Lautmuster von ihren genetisch fixierten Bedeutungen wäre nun der Aufwand, zwei verschiedene "Sprachen" für die Geschlechter aufrechtzuerhalten,

immer größer geworden. Da die sprechenden Hominiden aber nicht ganz auf eine zuverlässige akustische Signalisierung des Geschlechts verzichten konnten, mußte die Natur eine neue Möglichkeit finden, der (inzwischen weitgehend kulturell gewordenen) Sprache einen genetisch fixierten Dimorphismus aufzuprägen. Der Kehlkopf eignete sich insofern hierfür, als die Grundtonhöhe weitgehend unabhängig von den für die Semantik hauptsächlich maßgeblichen zeitlichen Artikulationsmustern variiert werden kann.

Welchen Stellenwert nimmt nun der zusätzlich beobachtbare Dimorphismus oberhalb 1000 Hz ein? Hierfür gibt es verschiedene Erklärungsmöglichkeiten:

- (1) Er ist eine zwangsläufige Folge allgemeiner anatomischer Geschlechterunterschiede
- (2) Er ist auf verhaltensphysiologisch zu erklärende Ursachen zurückzuführen, das heißt, auf neuromuskuläre Strukturen. Diese Strukturen können nun einerseits
 - (2a) kulturell erworben sein (im Sinne unterschiedlicher "Dialekte" der Geschlechter) oder
 - (2b) sie können genetisch fixierte Muster darstellen, die eventuell noch ein Relikt aus der präkulturellen Phase der phylogenetischen Sprachentwicklung darstellen.

Für alle drei Modelle gibt es Argumente; dennoch lassen sich gewisse Eingrenzungen machen:

Daß (1) nicht als vollständige Erklärung ausreicht, kann man daran sehen, daß die Korrelationen zwischen Konstitutionsmorphologie und Spektralverlauf im Langzeitspektrum, abgesehen von der Körpergröße und ihren Korrelaten im Körperbau, extrem niedrig sind, die (für den Geschlechtsdimorphismus nicht unmaßgebliche) Körpergröße sich aber anders im Spektrum ausdrückt als der Geschlechtsdimorphismus (vergleiche Abbildung 27, Seite 76, mit Abbildung 47, Seite 121).

Nimmt man (2a) als maßgebliche Erklärung an, so wäre zu erwarten, daß sich der Geschlechtsdimorphismus mehr in vom Sprecher erlernbaren Strukturen (zum Beispiel im Sprechrhythmus) ausdrückt als in weniger zugänglichen. Die gute Einsetzbarkeit des Langzeitspektrums zur Sprecheridentifizierung, die weitgehende Unabhängigkeit von einfachen willentlichen Einflußnahmen wie Variation von Lautstärke und Tonhöhe, der erhebliche Aufwand des Imitators, seine normale Variabilität zu verlassen und sein Unvermögen, in die objektive Variabilität des Imitierten zu gelangen, sowie die weitgehende Text- und Sprachunabhängigkeit sprechen eher gegen dieses Denkmodell. Bezüglich des beobachtbaren Geschlechtsdimorphismus im Sprechrhythmus wäre es eher in Erwägung zu ziehen.

Gegen (2b) könnte man vorbringen, daß es sich um eine spekulative Hypothese handelt, da man ja keine Anhaltspunkte für den zeitlichen Ablauf dieser phylogenetischen Entwicklung hat; Zwischenformen lassen sich keine beobachten und für die

Verhaltensphysiologie der Feinmotorik gibt es keine Fossilien. Immerhin gibt es keinen zwingenden Grund, (2b) zu verwerfen. Als einen, wenn auch schwachen, Hinweis, der für (2b) spricht, kann man anführen, daß der Geschlechtsdimorphismus im Langzeitspektrum oberhalb 1000 Hz auch in der ontogenetischen Entwicklung früher auftaucht als der (nach dieser Hypothese phylogenetisch jüngere) Dimorphismus in der Grundfrequenz, wie die Arbeiten von Bennett und Weinberg (1979) und Ingrisano et al. (1980) vermuten lassen. Die Autoren zeigten, daß das Geschlecht kleiner Kinder auditiv weit über dem Zufall erkannt wird. Da die Grundfrequenz in diesem Alter noch keine signifikanten Geschlechterunterschiede aufweist, liegt der von den Zuhörern erkannte Dimorphismus vermutlich in höheren Frequenzbereichen.

Es wäre dann eine interessante Tatsache, daß hier ein morphologisch – anatomisch erklärbares Merkmal (der Dimorphismus im Kehlkopf) phylogenetisch jünger ist als ein eher verhaltensphysiologisch erklärbares.

Zur Aufklärung des Mechanismus, über den in den höheren Frequenzbereichen das Geschlecht signalisiert wird, dürfte allerdings das Langzeitspektrum weniger geeignet sein. Man müßte als nächsten Schritt betrachten, wie der Dimorphismus auf der Betrachtungsebene einzelner Sprachlaute aussieht und welche Geschlechterunterschiede sich in dem den Sprachlauten zugrundeliegenden Artikulationsverhalten zeigen. Es wäre sinnvoll, diese Untersuchung nicht nur an Erwachsenen durchzuführen sondern auch an Kindern, eventuell sogar an den ersten Plapperlauten im Kleinkindalter, da sich hier ein eventuell genetisch fixierter Dimorphismus noch weitgehend unbeeinflusst von sprachkulturellen Einflüssen zeigen könnte. (Nach Enstrom 1982 sind sich die ersten Plapperlaute in verschiedenen Sprachgruppen wesentlich ähnlicher als die entsprechenden Laute in den Erwachsenensprachen, was mehr für "nature" als für "nurture" spricht)

5.2.4. Auditive Stimmbeurteilung und akustische Korrelate

Die Fragebögen zur auditiven Stimmbeurteilung wurden im wesentlichen nach Schneewind (1966) erstellt. Schneewind arbeitete allerdings mit einer ganz anderen Versuchsanordnung und hatte eine andere Fragestellung; daher sind seine Ergebnisse mit den hier dargestellten nur bedingt vergleichbar.

Eine Parallele ergibt sich in der Faktorenanalyse der Attribute: Auch bei Schneewind beinhaltet der erste Faktor, der auch den größten Varianzanteil beschreibt, allgemein emotional positiv wertende Eigenschaften (siehe Tabelle 4, Seite 101). Die übrigen Faktoren sind nicht vergleichbar, da Schneewind eine Varimax – Rotation des Faktorraums

durchführte. Obwohl diesem ersten Faktor in den Ohren der Zuhörer die größte Bedeutung zukommt (40% Varianzanteil), drückt er sich in akustischen Parametern nicht sehr stark aus, im Gegensatz zu Attributen wie "tief", "hell" oder "langsam".

Die Stimmselfbeurteilung, wie sie hier durchgeführt wurde (Vorspielen der eigenen Stimme über Kopfhörer) liefert bei den meisten Attributen Ergebnisse, die durchaus mit einer Fremdbeurteilung vergleichbar sind; deutlich schlechte Übereinstimmungen mit den Fremdbeurteilungsdaten waren nur bei den Begriffen "ungezwungen", "dünn", "weich" und "natürlich" zu beobachten.

Die Tatsache, daß das Mittel der 5 Beurteiler wesentlich besser mit der Selbstbeurteilung korreliert als die Daten der einzelnen Beurteiler (siehe Tabelle 3), sagt, daß durch die Mittelung über 5 Beurteiler die auf die Sprechproben und den Sprecher bezogenen Aussagen wesentlich zuverlässiger werden.

Bezüglich akustischer Korrelate auditiver Parameter und einer Computer-Schätzung der Attribute damit, kann folgendes gesagt werden (siehe auch Text Seite 110/111):

Eigenschaften, die sich auf Natürlichkeit, Sympathie oder Textverständlichkeit beziehen, lassen sich schlechter mit objektiven Methoden beschreiben als Eigenschaften, die sich auf Tonhöhe, Klang oder Sprechdynamik beziehen.

Es zeigte sich auch, daß die objektive Schätzung umso besser war, je homogener die 5 Beurteiler ankreuzten, mit Ausnahme der Begriffe "natürlich" und "deutlich".

5.2.5. Nonverbale Kommunikation von Sprechereigenschaften

Die Arbeiten verschiedener Autoren weisen darauf hin, daß es für den Zuhörer möglich ist, aus Sprechproben unbekannter Sprecher Eigenschaften der Sprecher zu erfassen. Solche Eigenschaften können psychologischer, soziologischer oder konstitutioneller Art sein.

Nach Scherer (1978) gibt es sogenannte "inference rules", die es dem Zuhörer ermöglichen, aufgrund von Vorerfahrungen, ähnlich wie es der Computer mit der Aufstellung von Regressionsfunktionen macht, Sprechereigenschaften zu schätzen.

Hier wurde ein Modell aufgestellt, das zum Testen solcher Zusammenhänge geeignet ist (siehe Abbildung 46, Seite 115). Insbesondere sollte dabei untersucht werden, welche objektiv meßbaren Aspekte des Sprachsignals mit der jeweiligen Sprechereigenschaft (z. B. Körpergröße) und mit ihren auditiven Entsprechungen (z. B. das Attribut "groß") korrelieren. Das Modell wurde auf die Merkmale Lebensalter, 2 Persönlichkeitsdimensionen, Körpergröße und auf die ersten beiden Hauptkomponenten aus 27 anthropometrischen Merkmalen angewendet. Signifikante Zusammenhänge waren nur

beim Lebensalter (im Bereich von 20–30 Jahren), der Körpergröße und der (mit der Körpergröße stark korrelierenden) ersten anthropometrischen Hauptkomponente festzustellen. Das Lebensalter wurde von den Zuhörern wesentlich besser geschätzt als vom Computer; bezüglich Körpergröße waren die objektive und die subjektive Schätzung ungefähr gleich gut. Die Tatsache, daß die Attributionen zum Teil stärker mit den aus objektiven akustischen Parametern gewonnenen Regressions-Schätzwerten korrelieren als mit den tatsächlichen Eigenschaften des Sprechers, weist darauf hin, daß sich der Zuhörer teilweise dieselbe Gesetzmäßigkeit für seine Schätzung zunutze macht wie der Computer.

Bezüglich Körpergröße kann die Beobachtung von Majewski et al. (1972) bestätigt werden, daß keine Korrelation zwischen Körpergröße und Sprechtonhöhe besteht; dennoch wird eine solche von den Zuhörern angenommen, woraus natürlich eine systematische Falscheinschätzung resultiert.

6. Zusammenfassung

Das Ziel der Arbeit ist, die menschliche Sprechstimme im Hinblick auf sprecherbezogene Eigenschaften zu untersuchen und damit einen Beitrag zur nonverbalen Kommunikation im Sprachsignal zu leisten.

Zunächst wird in einem Literaturüberblick eine Bestandaufnahme vorgenommen, die zeigen soll, was an Arbeiten auf diesem Gebiet bereits existiert. Dieser Überblick gliedert sich nach den Wissenschaftsdisziplinen Anthropologie, Sprach- und Informationswissenschaften, Psychologie und Medizin.

Es wurde dann eine Methode entwickelt, aus Sprechproben von ca. 20 Sekunden Dauer mit Hilfe der digitalen Signalverarbeitung einen Satz von etwa 300 Einzelparametern zu extrahieren; die Parameter beinhalten das Langzeitspektrum (mittleres Klangspektrum im Frequenzbereich von 80 bis 8000 Hz mit einer Frequenzauflösung von 80 Hz), das Langzeitspektrum im Frequenzbereich unterhalb 1000 Hz mit einer Auflösung von 10 Hz und das Langzeit-Rhythmusspektrum des Intensitätsverlaufs (im Bereich von ca. $1/3 - 6$ Hz, mit und ohne lineare Zeitkompression zur Adaption an die Sprechgeschwindigkeit) und des Grundfrequenzverlaufs (mit linearer Zeitkompression). Für die Aufzeichnung des Grundfrequenzverlaufs wurde ein auf die Fragestellung zugeschnittenes Verfahren neu konzipiert, das ermöglicht, ohne manuelle Eingriffe stetige Verlaufskurven längerer Sprechproben zu erhalten.

Damit auch bei begrenztem Stichprobenumfang die Anwendung multivariater statistischer Methoden möglich ist, mußten die Daten noch reduziert werden. Hierfür wurde im wesentlichen die Hauptkomponentenanalyse eingesetzt. Mit 20 Einzelparametern, also weniger als einem Zehntel der Ausgangsdaten, konnte ungefähr ein Varianzanteil von 80% beschrieben werden.

Mehrere Voruntersuchungen wurden durchgeführt. Dabei wurden der Einfluß technischer Fehlerquellen, die Textunabhängigkeit des Langzeitspektrums, der Einfluß von Sprechlautstärke und Tonhöhe auf das Langzeitspektrum und das Langzeitspektrum eines guten Stimmimitators untersucht; es wurden erste Hinweise auf einen Geschlechtsdimorphismus im Langzeitspektrum aufgezeigt und kasuistische Beobachtungen zur Sprachunabhängigkeit und zur Zwillingsähnlichkeit beschrieben.

Die Ergebnisse können folgendermaßen zusammengefaßt werden:

Das Langzeitspektrum stellt einen weitgehend text- und sprachunabhängigen Parametersatz dar; willentliche Beeinflussung von Sprechlautstärke und Tonhöhe bewirken zwar spektrale Veränderungen, gemessen an der interindividuellen Variabilität sind diese

jedoch gering, so daß zum Beispiel eine Sprecheridentifizierung noch weitgehend möglich ist.

Weitreichendere Verstellungen der Stimme, wie sie beim Imitieren anderer Sprecher erforderlich sind, bewirken, daß ein Sprecher seine normale Variabilität im Langzeitspektrum verläßt; allerdings gelingt es ihm meist nicht, in den intraindividuellen Streuungsbereich des Imitierten zu gelangen, obwohl sich der menschliche Zuhörer durchaus täuschen läßt. Insbesondere die mittlere Struktur des Langzeitspektrums (nach Glättung und Eliminierung des Grobverlaufs) erwies sich als relativ Sprecher-stabile Eigenschaft.

An einer größeren Stichprobe von Studenten wurde dann eine Untersuchung durchgeführt zur Beschreibung korrelativer Zusammenhänge zwischen akustischen Parametern aus dem Sprachsignal, auditivem Eindruck bezüglich Stimme und Sprecher, anthropometrischen Merkmalen und zwei Persönlichkeitsdimensionen des Sprechers.

Es wurden von 117 Versuchspersonen Klang- und Rhythmusspektren sowie sekundäre Parameter daraus berechnet. Zusätzlich wurden 27 anthropometrische Maße genommen, ein Persönlichkeits- und ein Stimmselfstbeurteilungsfragebogen ausgefüllt.

Sämtliche Sprechproben wurden fünf Phonetikstudenten vorgespielt, die jeweils auf einem Fragebogen den Sprecher in Bezug auf seine Stimme und Sprechweise beurteilen sollten, auf einem anderen in Bezug auf Aussehen (z. B. Körpergröße) und Persönlichkeit (z. B. Extrovertiertheit).

Die statistische Auswertung umfaßt folgende Punkte:

(1) Darstellung des Geschlechtsdimorphismus

Zusätzlich zu dem allgemein bekannten Geschlechtsdimorphismus in der mittleren Sprechtonhöhe finden sich Geschlechterunterschiede in höheren Frequenzbereichen, insbesondere im Bereich zwischen 3000 und 4000 Hz, sowie im Sprechrhythmus. Bei Verwendung aller akustischen Parameter läßt sich das Geschlecht unbekannter Sprecher in 99% aller Fälle richtig ermitteln; aus dem Frequenzbereich oberhalb 1000 Hz des Langzeitspektrums beträgt die Erkennungsrate noch über 90%; der Sprechrhythmus allein führt zu einer Erkennungsrate von über 70%. Die auditive Geschlechterkennungsrate liegt bei 98% und ist somit nicht höher als die durch den Computer.

Die Beobachtungen zum Geschlechtsdimorphismus werden im Zusammenhang mit der phylogenetischen Entwicklung des Menschen diskutiert; es wird eine Hypothese aufgestellt, nach der der Dimorphismus in höheren Frequenzbereichen phylogenetisch älter ist als der in der Grundfrequenz, obwohl ersterer eher verhaltensphysiologisch zu verstehen ist als daß er auf anatomische Geschlechterunterschiede zurückgeführt werden könnte. Diese Hypothese wird im Zusammenhang mit der Literatur diskutiert; es werden

Hinweise gebracht, die zur Prüfung beitragen können.

(2) Akustische Korrelate auditiver Parameter

Es wurde versucht, auditive Parameter, zum Beispiel die Ausprägung des Attributs "deutlich", objektiv akustisch zu beschreiben. Bei einigen Attributen war dies durchaus möglich, bei anderen schlechter: Eigenschaften, die sich auf Natürlichkeit, Sympathie oder Textverständlichkeit beziehen, ließen sich schlechter mit dem Computer schätzen als Eigenschaften, die sich auf Tonhöhe, Klang oder Sprechdynamik beziehen.

Die Stimmselfstbeurteilung brachte bei der gegebenen Versuchsanordnung (Vorspielen der eigenen Stimme über Kopfhörer) bei den meisten Attributen Ergebnisse, die mit denen einer Fremdbeurteilung vergleichbar sind. Lediglich bei Eigenschaften, die sich auf die Natürlichkeit der Stimme beziehen, besteht keine Korrelation zwischen Selbst- und Fremdbeurteilung.

(3) Nonverbale Kommunikation von Sprechereigenschaften

Es wurde ein Modell konstruiert, das, ähnlich dem Brunswick'schen Linsenmodell, aufzeigen soll ob und wie Eigenschaften des Sprechers nonverbal mit der Sprache signalisiert werden. Das Modell wurde auf Lebensalter, Konstitutions- und Persönlichkeitsparameter angewendet; positive Korrelationen zeigten sich nur beim Lebensalter und bei absoluten Größenmerkmalen, nicht bei Formmerkmalen und den untersuchten Persönlichkeitsdimensionen.

Das der vorliegenden Arbeit zugrundeliegende Datenmaterial liegt zur Einsicht im Institut für Anthropologie und Humangenetik bzw. im Leibnitz - Rechenzentrum.

Bibliographie

Die folgende Bibliographie enthält zusätzlich zu den im Text zitierten Literaturstellen eine umfangreiche Sammlung an Zitaten, die im weiteren thematischen Zusammenhang mit der vorliegenden Arbeit stehen. Sie wurde vollständig abgedruckt, da sie jemandem, der sich in die Thematik neu einarbeiten möchte, vielleicht von Nutzen ist. Da die Zitate zum Teil aus anderen Literaturverzeichnissen übernommen wurden, sind die Angaben nicht alle einheitlich (bei manchen Büchern fehlt zum Beispiel die Angabe über den Verlag).

Hinter jeder Literaturstelle sind in eckigen Klammern bis zu drei Zahlen aufgeführt, die den Kategorien aus dem Suchindex entsprechen. Im Index selbst sind zu jeder Kategorie die Nummern der entsprechenden Literaturangaben aufgelistet.

Der Index gliedert sich folgendermaßen (die Lücken in der Nummerierung sind beabsichtigt, damit das System gegebenenfalls erweitert werden kann):

A) Biological aspects

- 9 phylogeny, primatology
- 10 anatomy
- 11 larynx
- 12 physiology, EEG, EMG, ...
- 13 genetics
- 14 age
- 15 child development
- 16 sex
- 17 race
- 18 twins
- 19 constitution

B) Medical aspects

- 20 voice quality
- 21 psychiatry
- 22 drug influences
- 23 menstruation, birth ...
- 24 castrato voice
- 25 other medical aspects

C) Speech and language sciences

- 30 languages
- 31 speech analysis
- 32 automatic speech recognition
- 33 speech synthesis
- 34 speaker adaptation

D) Psychology and sociology

- 40 psychology general
- 41 personality
- 42 behavioural sciences
- 43 social aspects
- 44 emotions, psychology of expression

E) Acoustical analysis

- 50 intensity, pitch and F0
- 51 spectrographic methods
- 52 average speech spectrum
- 53 general voice acoustics, models of speech production
- 54 prosody, speech rhythm
- 55 auditive analysis

F) Cetera

- 60 signal processing
- 61 statistics
- 62 experimental design
- 63 literature
- 64 monographs, collections
- 65 speaker recognition
- 66 handwriting

- 1 Abercrombie, D. et al. 1964: In Honour of Daniel Jones, Longman, London [64]
- 2 Addington, D. W. 1968: The relationship of selected vocal characteristics to personality perception, *Speech Monographs* 35, 492–503 [41]
- 3 Addington, D. W. 1968: Voice and the perception of personality: an experimental study, *Oklahoma State University Monographs, Social Sciences Series* 15 [41]
- 4 Aitken, P. G. 1981: Cortical control of conditioned and spontaneous vocal behaviour in Rhesus monkeys, *Brain and Language* 13, 171–184 [9 42]
- 5 Allen, G. D. (chairman) 1982: Session K: Speech communication 3, developmental and cross-linguistic phonetics, *J. acoust. soc. Amer.* 71, S21–S23 [15 31]
- 6 Allen, G. D. 1982: Cross-language study of children's prosodic development 2: Syllabic duration, F0 and intensity in 2 years olds' words, *J. acoust. soc. Amer.* 71, S21 [15]
- 7 Allen, G. D. 1983: Some suprasegmental contours in French two-year-old children's speech, *Phonetica* 40, 269–292 [15]
- 8 Allport, G. W. and Cantril, H. 1934: Judging personality from voice, *J. soc. psychol.* 5, 37–55 [41]
- 9 Alpert, M. and Kurtzberg, R. L. 1963: Comparison of the spectra of twins, *J. acoust. soc. Amer.* 35, 1877 [15]
- 10 Altmann, B., Goydke, H. and Schlegel, D. 1981: Sprachlautspektren nach Eingliederung von Gaumenplatten, *Fol. phoniat.* 33, 261–274 [25]
- 11 Amorosa, H. 1982: The timing of speech and hand motor coordination in language delayed children, *J. acoust. soc. Amer.* 71, S22 [21 25]
- 12 Anthony, J. 1964: True model of the vocal tract, *J. acoust. soc. Amer.* 36, 1037 [10 53]
- 13 Ardener, E. (Ed.) 1970: Social anthropology and language, Association of social anthropologists monograph 10, Tavistock press, London [43]
- 14 Arndorfer, H. 1965: Zur Biophonetik des rhetorischen Ausdrucks, in: Zwirner, E: Proceedings of the 5th international congress of phonetic sciences, S. Karger, Basel, 207–210 [42 19]
- 15 Arndt, H. J. v. 1963: Untersuchungen über den Einfluß der Mandelausschaltung auf die menschliche Stimme, *Fol. phoniat.* 15, 99–146 [25]
- 16 Aronovitch, C. D. 1976: The voice of personality: stereotyped judgements and their relation to voice quality and sex of speaker, *J. of social psychology* 99, 207–220 [16 20 41]
- 17 Atal, B. S. 1972: Automatic speaker recognition based on pitch contours, *J. acoust. soc. Amer.* 52, 1687–1697 [65]
- 18 Atal, B. S. 1972: Determination of articulatory parameters from the speech signal, *J. acoust. soc. Amer.* 52, 182 [31]
- 19 Atal, B. S. 1972: Text-independent speaker recognition, *J. acoust. soc. Amer.* 52, 181 [65]
- 20 Atal, B. S. 1976: Automatic recognition of speakers from their voices, *Proc. IEEE* 64, 460–475 [65]
- 21 Atal, B. S. and Hanauer, S. L. 1971: Speech analysis and synthesis by linear prediction of the speech wave, *J. acoust. soc. Amer.* 50, 637–644 [31 33]
- 22 Atal, B. S. and Schröder, M. R. 1978: Linear prediction analysis of speech based on a pole-zero representation, *J. acoust. soc. Amer.* 64, 1310–1318 [60]
- 23 Austin, M. and Leeper, H. 1975: Basal pitch and frequency level variation in male and female children: A preliminary investigation, *J. communication disorders* 8, 309–316 [15 50]
- 24 Böhme, G. 1966: Störungen der Sprache, der Stimme und des Gehörs durch frühkindliche Hirnschädigungen. Klinische und experimentell phonetische Untersuchungen, Fischer, Jena [21 25 51]
- 25 Böhme, G. 1968: Intersexualität und Stimme, *Fol. phoniat.* 20, 417–427 [16 25]
- 26 Bamberg, G. and Baur, F. 1980: Statistik, Oldenbourg, München [61]
- 27 Baran, J. and Seymour, H. N. 1976: The influence of three phonological rules of black English on the discrimination of minimal word pairs, *J. speech hear. res.* 19, 467–474 [31 17]

- 28 Barth, E. 1911: Einführung in die Physiologie, Pathologie und Hygiene der menschlichen Stimme, Thieme, Leipzig [12 25]
- 29 Bauer, H. 1963: Die Beeinflussung der weiblichen Stimme durch androgene Hormone, *Fol. phoniat.* 15, 264–268 [23 16]
- 30 Becker, M. H. et al. 1964: Comparisons of some statistical distance measures for talker identification, *J. acoust. soc. Amer.* 36, 1988 [61 65]
- 31 Becker, R. W. 1966: Characteristics that determine speaker recognition, *Electronic system division ESD – TR – 66 – 636*, 1–66 [65]
- 32 Beckman, M. E. 1982: Effects of accent on vowel duration in Japanese, *J. acoust. soc. Amer.* 71, S23 [30]
- 33 Behrendt, W. and Strauch, G. 1965: Die Feinstruktur des menschlichen Stimmbandes in Abhängigkeit vom Lebensalter, *Arch. Ohr Nas Kehlkopfheilk.* 184, 510–520 [10 14]
- 34 Bell, C.G. et al. 1961: Reduction of speech spectra by analysis – by – synthesis techniques, *J. acoust. soc. Amer.* 33, 1725–1736 [31 33]
- 35 Bellet, G., Contini, M. and Boe, L. J. 1982: Normalisation temporelle et frequentielle de F0 intra et inter individuelle, *Bull. de l'institut de phonetique de Grenoble* 10/11, 265–283 [31 32]
- 36 Benjamin, B. 1982: Phonological performance in gerontological speech, *J. psycholinguist. res.* 11, 159–167 [14 51]
- 37 Benjamin, B. J. 1981: Frequency variability in the aged voice, *J. Gerontology* 36, 722–726 [14]
- 38 Bennett, S. 1983: A 3-year longitudinal study of school-aged children's fundamental frequencies, *J. speech hear. res.* 26, 137–142 [15 50]
- 39 Bennett, S. and Weinberg, B. 1979: Acoustic correlates of perceived sexual identity in preadolescent children's voices, *J. acoust. soc. Amer.* 66, 989–1000 [15 16]
- 40 Berendes, J. 1968: Die Verantwortlichkeit des Arztes bei der Anwendung anaboler Steroide im Hinblick auf die Stimme, *Fol. phoniat.* 20, 379–386 [23 16]
- 41 Berendes, Link, Zöllner 1983: Kehlkopf 2, Hals – Nasen – Ohren – Heilkunde in Praxis und Klinik Bd. 4, Teil 2, Thieme, Stuttgart [11 25]
- 42 Bernstein, F. 1923: Zur Statistik der sekundären Geschlechtsmerkmale beim Menschen, *Nachrichten von der königlichen Gesellschaft der Wissenschaften zu Göttingen* 2, Weidmann, Berlin [13 16]
- 43 Besch, W., Knoop, U., Putschke, W., Wiegand, H. E. 1982: *Dialektologie, ein Handbuch zur deutschen und allgemeinen Dialektforschung*, de Gruyter, New York [30]
- 44 Beutel, P. and Schubö, W. 1983: SPSS9, Statistik – Programmpaket für die Sozialwissenschaften, nach N. H. Nie und C. H. Hull, Fischer, Stuttgart [61]
- 45 Black, J. W. 1961: Relationships among fundamental frequency, vocal sound pressure and rate of speaking, *Language and speech* 4, 196–199 [50 54]
- 46 Black, J. W. et al. 1973: Reply to speaker identification by speech spectrograms: some further observations, *J. acoust. soc. Amer.* 54, 535–537 [51 65]
- 47 Bladon, A. 1982: Arguments against formants in the auditory representation of speech, in: Carlson, R. and Granström, B. (Ed): *The representation of speech in the peripheral auditory system*, Elsevier Biomedical Press, Amsterdam, 95–102 [31]
- 48 Bloch, P. and Bloch, M. 1983: Voice identity, *Fol. phoniat.* 35, 107 [40 41]
- 49 Blumstein, S. E. and Stevens, K. N. 1979: Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants, *J. acoust. soc. Amer.* 66, 1001–1017 [31 51]
- 50 Blumstein, S. E. and Stevens, K. N. 1980: Perceptual invariance and onset spectra for stop consonants in different vowel environments, *J. acoust. soc. Amer.* 67, 648–662 [31]
- 51 Boe, L. J. and Abry, C. 1980: Caracteristiques individuelles et phonetiques, *Bull. inst. phonet. Grenoble* 9, 1–16 [31 34]
- 52 Boe, L. J. and Bellet, G. 1979: Mesure et statistique de la frequence laryngienne. Appareillage et programmes, *Bull. inst. phonet. Grenoble* 8, 23–34 [50 54]
- 53 Boe, L. J., Contini, M. and Rakotofringa, H. 1975: Etude statistique de la frequence laryngienne. Applications a l'analyse et la synthese des faits prosodiques du francais, *Phonetica* 32, 1–23 [50 54 16]

- 54 Boe, L. J., Pascal, D. and Raymond, J. 1982(?): Les caracteristiques individuelles de la parole, Bibliographie (a paraître) [63]
- 55 Boehme, G. 1972: Untersuchungsmethoden der Stimme und Sprache, Barth, Leipzig [25]
- 56 Boehme, G. 1978: Methoden zur Untersuchung der Sprache, des Sprechens und der Stimme, Fischer, Stuttgart [64]
- 57 Boehme, G. 1982: Therapie der Sprach-, Sprech- und Stimmstörungen, Fischer, Stuttgart [25]
- 58 Bogner, R. E. 1981: On talker verification via orthogonal parameters, IEEE - ASSP 29, 1 - 12 [65]
- 59 Bogner, R. E. 1981: On talker verification via orthogonal parameters, IEEE - ASSP 29, 1 - 12 [65]
- 60 Bolt, R. H. et al. 1973: Speaker identification by speech spectrograms. Some further observations, J. acoust. soc. Amer. 54, 531 - 534 [65 51]
- 61 Bolt, R. H. et al. 1979: Speaker identification by speech spectrogram, J. acoust. soc. Amer. 47, 597 - 613 [65 51]
- 62 Bonaventura, M. 1935: Ausdruck der Persönlichkeit in der Sprechstimme und im Photogramm, Arch. ges. Psychol. 94, 501 - 570 [41]
- 63 Bond, R. N. and Feldstein, S. 1982: Acoustical correlates of the perception of speech rate: An experimental investigation, J. psycholing. res. 11, 539 - 557 [54 55]
- 64 Borucki, H. 1973: Einführung in die Akustik, Mannheim [53]
- 65 Boulogne, M. 1979: Une nouvelle methode de detection du voisement: Etude et evaluation, Bull. inst. phonet. Grenoble 8, 1 - 21 [31]
- 66 Bracken, H. v. 1939: Das Schreibtempo von Zwillingen und die sozialpsychologischen Fehlerquellen der Zwillingsforschung, Z. menschl. Vererb. u. Konst. 23, 278 - 298 [18 66]
- 67 Bracken, H. v. 1939: Untersuchungen an Zwillingen über die qualitativen und quantitativen Merkmale des Schreibdrucks, Z. angew. Psychol. 58, 367 - 384 [18 66]
- 68 Bracken, H. v., Pongs, L. and Riedel, H. 1949: Ein neues Gerät zum Registrieren des Schreibdrucks, Abh. d. Braunschweigischen Wiss. Ges., Braunschweig, 126 - 131 [66]
- 69 Bricker, P. D. and Pruzansky, S. 1965: Temporal effects in talker identification, J. acoust. soc. Amer. 38, 924 [55 65]
- 70 Bricker, P. D. and Pruzansky, S. 1966: Effects of stimulus content and duration on talker identification, J. acoust. soc. Amer. 40, 1441 - 1449 [55 65]
- 71 Bricker, P. D. et al. 1971: Statistical techniques for talker identification, Bell system technical journal 50, 1427 - 1454 [61 65]
- 72 Broad, D. J. 1968: Some physiological parameters for prosodic description, Speech communication research laboratory, Santa Barbara [12 54]
- 73 Broad, D. J. 1970: Research on the prosodic aspects of speech, Speech communications research lab., Santa Barbara [54]
- 74 Brodnitz, F. S. 1962: The holistic study of the voice, Quart. J. of speech 48, 280 - 284 [20 40]
- 75 Brodnitz, F. S. 1975: The age of the castrato voice, J. speech hear. dis. 40, 291 - 295 [24]
- 76 Brosnahan, L. F. 1959: Genetics and the "th" - sounds, Phonetica 4, 183 - 190 [13]
- 77 Brosnahan, L. F. 1961: The sounds of language - an inquiry into the role of genetic factors in the development of sound systems, Heffer & sons, Cambridge [13]
- 78 Brouwer, P. 1977: Signalanalyse für eine sprecherunabhängige Spracherkennung, Diss. TU Braunschweig [32]
- 79 Brown, B. L. 1969: A social psychology of variations in French Canadian speech styles, unpubl. diss., Mc Gill Univ. (Microfilm copy in the national lib. in Ottawa) [43 30]
- 80 Brown, B. L. 1976: A cross cultural study of social status markers in speech, Can. j. of behavioral science 8, 39 - 55 [43]
- 81 Brown, B. L., Strong, W. J. and Rencher, A. C. 1973: Perceptions of personality from speech: effects of manipulations of acoustical parameters, J. acoust. soc. Amer. 54, 29 - 35 [41 55]

- 106 Coleman, R. O. 1973: Speaker identification in the absence of inter-subject differences in the glottal source characteristics, *J. acoust. soc. Amer.* 53, 1741-1743 [65]
- 107 Compton, A. J. 1963: Effects of filtering and vocal duration upon the identification of speakers aurally, *J. acoust. soc. Amer.* 35, 1748-1752 [65]
- 108 Constanzo, F. S., Markel, N. N. and Constanzo, P. R. 1969: Voice quality profile and perceived emotion, *J. counseling psychol.* 16, 267-270 [20 44]
- 109 Corsi, P. 1978: Introduction à la reconnaissance du locuteur: 1. - contexte de l'étude et problèmes fondamentaux, *Bull. inst. phonet. Grenoble* 7, 144-169 [65]
- 110 Cort, S. and Murry, T. 1972: Aural identification of children's voices, *J. acoust. soc. Amer.* 51, 131 [15 65]
- 111 Cremer, L. 1979: Physikalische Analogien zur Faktorenanalyse, *Acustica* 42, 1-17 [61]
- 112 Crystal, D. 1969: Prosodic systems and intonation in English, Cambridge University Press, Cambridge [50 54]
- 113 Crystal, D. 1970: Prosodic and paralinguistic correlates of social categories, in: Ardener, E. (Ed.): *Social anthropology and language*, Assoc. social anthropol. monogr. 10, Tavistock, London [43 54]
- 114 Crystal, D. and Davy, D. 1969: *Investigating English style*, Longmans, London [30]
- 115 Curry, E. 1940: The pitch characteristics of the adolescent male voice, *Speech monogr.* 7, 48-62 [15 50]
- 116 Cutler, E. et al. 1972: The evidentiary value of spectrographic voice identification, *J. crim. law, criminol. and police science* 63, 343-355 [51 65]
- 117 Damste, P. H. 1964: Virilization of the voice due to anabolic steroids, *Fol. phoniat.* 16, 10-18 [16 22]
- 118 Damste, P. H. 1967: Voice change in adult women caused by virilizing agents, *J. speech hear. dis.* 32, 126-132 [16 22]
- 119 Danhauer, J. L. and Singh, S. 1975: A multidimensional scaling analysis of phonemic responses from hard of hearing and deaf subjects of three languages, *Lang. speech* 18, 42-64 [25]
- 120 Darlington, C. D. and Mather, K. 1949: *Elements of genetics*, Allen and Unwin, London [13]
- 121 Das, S. K. and Mohn, W. S. 1971: A scheme for speech processing in automatic speaker verification, *IEEE - AU* 19, 32-43 [65]
- 122 Davitz, J. R. and Davitz, L. J. 1959: Correlates of accuracy in the communication of feelings, *J. communication* 9, 110-117 [44 55]
- 123 Davitz, J. R. and Davitz, L. J. 1959: The communication of feelings by content-free speech, *J. communication* 9, 6-13 [44 55]
- 124 Deal, L. V. and Oger, H. J. 1983: Pitch and vocal pleasantness in the aging process, *Fol. phoniat.* 35, 119 [14 50 55]
- 125 Deinse, J. B. v., Frateur, L. and Keizer, J. 1974: Problems of the singing voice, *Fol. phoniat.* 26, 428-434 [25]
- 126 Deuster, C. v. 1977: Irreversible Stimmstörung in der Schwangerschaft, *HNO* 25, 430-432 [23 50]
- 127 Diehl, C. F. 1960: Voice and personality: An evaluation, in: Barbara, D. A. (Ed): *Psychological and psychiatric aspects of speech and hearing*, Springfield, Ill. [41]
- 128 Diehl, C. F., White, R. and Burle, K. 1959: Voice quality and anxiety, *J. speech hear. res.* 2, 282-285 [20 44]
- 129 Diller, T. C. 1981: Speaker-independent word recognition using sex-dependent clustering, *ICASSP* 81, Atlanta, 180-183 [32]
- 130 Diller, T. C. and Siebenand, J. F. 1981: Speaker-independent word recognition using sex-dependent clustering, *ICASSP* 81, Proc. IEEE, 180-183, Phys. Abstr. 68679 [32]
- 131 Doddington, G. R. 1971: A method of speaker verification, *J. acoust. soc. Amer.* 49, 139 [51 60 65]
- 132 Doddington, G. R. 1976: Personal identity verification using voice, Paper on: *Electro* 76, Boston [65]

- 133 Doddington, G. R. and Schalk, T. B. 1981: Speech recognition: turning theory to practice, *IEEE Spectrum* 18/9, 26–33 [32]
- 134 Doddington, G. R., Hydrick, B. and Beek, B. 1973: Some results on speaker verification using amplitude spectra, 86th meeting of the acoust. soc. Amer. [51 60 65]
- 135 Doherty, E. T. 1976: An evaluation of selected acoustic parameters for use in speaker identification, *J. Phonetics* 4, 312–326 [52 65]
- 136 Dommelen, W. v. 1982: A contrastive investigation of vowel duration in German and Dutch, *Phonetica* 39, 23–35 [30 31]
- 137 Dordain, M. 1972: Etude statistique de l'influence des contraceptifs hormonaux sur la voix, *Fol. phoniatic.* 24, 86–96 [16 22]
- 138 Duda, R. O. and Hart, P. E. 1973: Pattern classification and scene analysis, Wiley, New York [61]
- 139 Duffy, R. 1958: The vocal pitch characteristics of eleven, thirteen and fifteen–years–old female speakers, unpubl. diss., State univ. of Iowa [15 50]
- 140 Duffy, R. J. 1970: Fundamental frequency characteristics of adolescent females, *Lang. speech* 13, 14–24 [15 50]
- 141 Duifhuis, H. and Willems, L. F. 1982: Measurement of pitch in speech: An implementation of Goldstein's theory of pitch perception, *J. acoust. soc. Amer.* 71, 679–688 [50]
- 142 Duncan, M. H. 1945: An experimental study of some of the relationship between voice and personality among students of speech, *Speech monogr.* 12, 47–60 [41]
- 143 Dusenbury, D. and Knower, F. H. 1939: Experimental studies of the symbolism of action and voice 2: A study of the specificity of meaning in abstract tonal symbols, *Quart. j. speech* 25, 67–75 [44]
- 144 Eberle, A. 1979: Mutationsstimmstörung als Identifikationsproblem – ein Therapiekonzept, *Sprache – Stimme – Gehör* 3, 110–113 [15 25]
- 145 Eggert, D. 1974: Eysenck – Persönlichkeits – Inventar (EPI), Hogrefe, Göttingen [40]
- 146 Eguchi, S. and Hirsh, I. J. 1969: Development of speech sounds in children, *Acta Otolaryngologica* 257 (Suppl.) [14]
- 147 Eimas, P. D. 1985: Sprachwahrnehmung beim Säugling, *Spektrum der Wissenschaft*, März 85, 76–83 [15 55 62]
- 148 Eisenberg, P. and Zalowitz, E. 1938: Judgments of dominance–feeling from phonograph records of voice, *J. appl. psychol.* 22, 620–631 [41]
- 149 Ekman, P., Friesen, W. V. and Scherer, K. R. 1976: Body movement and voice pitch in deceptive interaction, *Semiotica* 16, 23–27 [42 50]
- 150 Ekman, P., Friesen, W. V. and Scherer, K. R. 1976: Body movement and voice pitch in deceptive interaction, *Semiotica* 16, 23–27 [44 50]
- 151 El Chafei, C. 1980: A real time speaker recognition system on mini computer, Carnahan conference, Lexington, KY [65]
- 152 Ellis, D. S. 1967: Speech and social status in America, *Social forces* 45, 431–437 [43]
- 153 Endress, W., Bambach, W. and Flosser, G. 1971: Voice spectrograms as a function of age, voice disguise and imitation, *J. acoust. soc. Amer.* 49, 1842–1848 [14 51 65]
- 154 Enke, W. 1940: Motorik und Psychomotorik, in: Just, G.: *Handbuch der Erbbiologie des Menschen* 2, Berlin [12 42]
- 155 Enstrom, D. H. 1982: Infant labial, apical and velar stop productions: A voice onset time analysis, *Phonetica* 39, 47–60 [15 54]
- 156 Enstrom, D. H. and Spörrli–Büttler, S. 1981: A voice onset time analysis of initial Swiss–German stops, *Fol. phoniatic.* 33, 137–150 [30]
- 157 Entwisle, D. and Greenberger, E. 1969: Racial differences in the language of grade school children, *Sociology of education* 42, 238–250 [17 30]
- 158 Ertel, S. 1969: Psychophonetik, Untersuchungen über Lautsymbolik und Motivation, *Verl. f. Psychologie Hogrefe, Göttingen* [40 31]
- 159 Essen, O. v. (Ed.) 1952: *Sprechmelodie als Ausdrucksgestaltung*, Buske, Hamburg [54 40]

- 160 Essen, O. v. 1951: Ein einfaches Maß für die Durchschlagskraft der Stimme, *FoL phoniat.* 3, 232 – 240 [20 55]
- 161 Essen, O. v. 1953: *Allgemeine und angewandte Phonetik*, Akad. Verl., Berlin [31 64]
- 162 Essen, O. v. 1961: *Mathematische Analyse periodischer Vorgänge in gemeinfälliger Darstellung*, Elwert, Marburg [60]
- 163 Eysenck, H. J. and Eysenck, S. B. G. 1964: *Manual of the Eysenck Personality Inventory*, University of London Press, London [41]
- 164 Fährmann, R. 1967: *Die Deutung des Sprechausdrucks: Studien zur Einführung in die Praxis der charakterologischen Stimm- und Sprechanalyse*, Bouvier, Bonn [40 55]
- 165 Fährmann, R. 1973: Die kriminalphonetische Expertise und der Sprechstimmensachverständige heute, *Kriminalistik* 27, 97 [40 55]
- 166 Fährmann, R. 1977: *Simm- und Sprechdiagnostik. Leitfaden der angewandten Sprechanalyse*, *Z. f. Menschenkunde* 41, 57 – 123 [41 55]
- 167 Fairbanks, G. 1940: Recent experimental investigations of vocal pitch in speech, *J. acoust. soc. Amer.* 11, 547 – 466 [15 44 50]
- 168 Fairbanks, G. 1942: An acoustical study of infant hunger wails, *Child develop.* 13, 227 – 232 [15 50]
- 169 Fairbanks, G. 1966: *Experimental phonetics. Selected articles*, Urbana, London [31 40 64]
- 170 Fairbanks, G. and Hoaglin, L. W. 1941: An experimental study of the durational characteristics of the voice during the expression of emotion, *Speech monogr.* 8, 85 – 90 [44 54]
- 171 Fairbanks, G. and Pronovost, W. 1939: An experimental study of the pitch characteristics of the voice during the expression of emotions, *Speech monogr.* 6, 87 – 104 [44 50]
- 172 Fairbanks, G., Herbert, E. and Hamrmond, J. 1949: An acoustical study of vocal pitch in seven – year – old girls, *Child develop.* 20, 71 – 78 [15 50]
- 173 Fairbanks, G., Wiley, F. and Lassman, L. 1949: An acoustical study of vocal pitch in seven – year – old boys, *Child develop.* 20, 63 – 71 [15 50]
- 174 Fant, G. 1962: *The acoustic theory of speech production*, Mouton, the Hague [53 64]
- 175 Fant, G. 1970: *The acoustic theory of speech production. With calculations based on X – ray studies of Russian articulations.* 2nd print, Mouton, the Hague [53 64]
- 176 Fant, G. 1973: *Speech sounds and features*, MIT press, Cambridge [53 64]
- 177 Fant, G. and Scully, C. (Ed.) 1977: *The larynx and language*, *Phonetica* 34 (spec. iss.) [11 31]
- 178 Fastl, H. 1976: Schallpegel und Lautstärke von Sprache, *Acustica* 35, 341 – 345 [50 55]
- 179 Fay, P. J. and Middleton, W. C. 1939: Judgment of Spranger personality types from the voice as transmitted over a public address system, *Charact. pers.* 8, 144 – 155 [41]
- 180 Fay, P. J. and Middleton, W. C. 1939: Judgment of occupation from from the voice as transmitted over a public address system and over radio, *J. appl. psychol.* 23, 586 – 601 [43]
- 181 Fay, P. J. and Middleton, W. C. 1940: Judgment of Kretschmerian body types from the voice as transmitted over a public address system, *J. soc. psychol.* 12, 151 – 162 [19]
- 182 Fay, P. J. and Middleton, W. C. 1940: The ability to judge sociability from the voice as transmitted over a public address system, *J. soc. psychol.* 13, 303 – 309 [41 43]
- 183 Fay, P. J. and Middleton, W. C. 1940: The ability to judge the tested or tired condition of a speaker from his voice as transmitted over a public address system, *J. appl. psychol.* 24, 645 – 650 [44]
- 184 Fay, P. J. and Middleton, W. C. 1942: Judgment of introversion from the transcribed voice, *Quart. j. speech* 28, 226 – 228 [41]
- 185 Fay, P. J. and Middleton, W. C. 1943: Judgment of leadership from the transcribed voice, *J. soc. psychol.* 17, 99 – 102 [41]

- 186 Fedders, B. and Schultz-Coulon, H. J. 1975: Ein Grundtonanalysator für die klinische und experimentelle Phoniatrie, *Acustica* 33, 17 [50]
- 187 Federico, A. et al. 1981: Recognition test for an open set of talkers represented by their acoustic parameters, Pp. 6., *Fondazione Ugo Bordonì, Rome* (Mar. 1981), N81-31446/0 (Nr. 2, 1982) [65]
- 188 Ferenbach, M. 1950: Probleme der Sprechanalyse, *Psychol. Rundsch.* 1, 162-168 [40 55]
- 189 Ferrari Disner, S. 1983: Vowel quality. The relation between universal and language-specific factors, *UCLA working papers in phonetics* 58 [31]
- 190 Fertig, B. L. 1983: A workshop in meditation and visualisation as applied to therapy with functional voice disorders: A case study, *Fol. phoniät.* 35, 122 [20 25]
- 191 Fink, B. R. and Demarest, R. J. 1978: *Laryngeal biomechanics*, Harvard univ. press, Cambridge, Mass. [11 64]
- 192 Finn, J. D. 1974: A general model for multivariate analysis, Holt, Rinehart & Winston, New York [61]
- 193 Fishman, B. v. et al. 1971: The effects of certain drugs on phonation, *J. speech hear. res.* 14, 301-306 [22]
- 194 Flach, M., Schwickardi, H. and Heidelberg, J. G. 1980: Die Wertigkeit menstrueller Stimmveränderungen bei Gesangsstudentinnen, *Fol. phoniät.* 32, 347-352 [23]
- 195 Flach, M., Schwickardi, H. and Simon, R. 1968: Welchen Einfluß haben Menstruation und Schwangerschaft auf die ausgebildete Singstimme, *Fol. phoniät.* 21, 199-210 [23]
- 196 Fonagy, I. 1958: Elektrophysiologische Beiträge zur Akzentfrage, *Phonetica* 2, 12-58 [54 12]
- 197 Fonagy, I. 1962: Mimik auf glottaler Ebene, *Phonetica* 8, 209-219 [11 42 44]
- 198 Foster, R. A. 1982: Voice recognition: a word about its future, *Computerworld* 16 (11a), 39-40 [65]
- 199 Frank, F. 1980: Erfahrungen mit Eignungs- und Tauglichkeitsuntersuchungen bei Sprech- und Stimmberufen in Österreich, *Sprache - Stimme - Gehör* 4, 112-116 [25]
- 200 Frank, F. and Donner, F. 1981: Stimmumfänge in Abhängigkeit von der Tageszeit (mit sonographischer Analyse), *Sprache - Stimme - Gehör* 5, 177-180 [20 12]
- 201 Frank, F. and Tesarek, L. 1980: Neue Erkenntnisse über Stimmumfänge bei sechsjährigen Kindern, *Sprache - Stimme - Gehör* 4, 147-149 [15]
- 202 Frank, F. et al. 1979: Neue Erkenntnisse über Stimmumfänge bei drei- bis fünfjährigen Kindern bzw. Kindergartenkindern, *Sprache - Stimme - Gehör* 3, 167-170 [15 50]
- 203 Freedman, D. G. 1961: The differentiation of identical and fraternal twins on the basis of filmed behaviour, 2nd int. conf. human genetics, Rome [18]
- 204 Frehse, R. et al. 1980: Automatic recognition of cooperative speakers, Pp. 168, *Philips Forschungslab., Hamburg*, N82-12319/1 (7/82) [65]
- 205 Frehse, R. et al. 1982: Automatic recognition of comparative speakers, Pp. 107, *Philips Forschungslab., Hamburg*, N82-12319/1 (7/82) [65]
- 206 Frender, R., Brown, B. L. and Lambert, W. E. 1970: The role of speech characteristics in scholastic success, *Can. j. behav. sc.* 2, 299-306 [40]
- 207 Friedhoff, A. J., Alpert, M. and Kurtzberg, R. L. 1962: An effect of emotion on the voice, *Nature* 193, 357-358 [44 51]
- 208 Friedhoff, A. J., Alpert, M. and Kurtzberg, R. L. 1964: Infra-content channels of vocal communication, in: *Disorders of communication*, Vol. 42, Res. publ., A. R. N. M. D. [44 51]
- 209 Furui, S. 1977: Analysis of temporal variation of talker dependent features, *Review of the ELectrical Communication Laboratories* 25, 231-244 [34 65]
- 210 Furui, S. 1981: Cepstral analysis technique for automatic speaker verification, *IEEE-ASSP* 29, 254-272 [65]
- 211 Furui, S. 1981: Comparison of speaker recognition methods using statistical features and dynamic features, *IEEE-ASSP* 29, 342-350 [65]
- 212 Furui, S. 1981: Comparison of speaker recognition methods using statistical features and dynamic features, *IEEE-ASSP* 29, 342-350 [65]

- 213 Furui, S., Itakura, F. and Saito, S. 1972: Talker recognition by long time averaged speech spectrum, *Electronics and communications in Japan* 55 - A, 54 - 61 [52 65]
- 214 Gackenholtz, S. 1979: Akustische Untersuchungen der Phonation während der Therapie von Rekurrensparesen, Diss. med., München [25]
- 215 Garvin, P. L. and Ladefoged, P. 1963: Speaker identification and message identification in speech recognition, *Phonetica* 9, 193 - 199 [65]
- 216 Gedda, L. et al. 1955: La voce dei gemelli, *Acta genet. med. gemellolog.* 2, 121 [18 55]
- 217 Gedda, L., Fiori-Ratti, L. and Bruno, G. 1960: La voix chez les jumeaux monozygotiques, *Fol. phoniat.* 12, 81 - 94 [18]
- 218 Geller, D. and Margolis, R. H. 1984: Magnitude estimation of loudness 1: Application to hearing aid selection, *J. speech hear. res.* 27, 20 - 27 [12 55]
- 219 Gerull, G. et al. 1977: Untersuchungen verschiedener Gruppen von Stimmstörungen durch statistische Spektralanalyse, *HNO* 25, 14 [25 51]
- 220 Giet, G. van der et al. 1980: Forensic speaker recognition system for police applications, *International conference: security through science and engeneering*, Berlin, 50 - 57 [65]
- 221 Giet, G. van der et al. 1980: Rechnergestützte Sprechererkennung für Anwendungen im Polizeibereich, 5. internationale Konferenz über spezielle Polizeitechnik, 233 - 252 [65]
- 222 Gilbert, H. R. and Weismer, G. G. 1974: The effects of smoking on the speaking fundamental frequency of adult women, *J psycholinguistic res.* 3, 225 - 231 [25 50]
- 223 Giles, H., Bourhis, R. Y. and Davis, A. P. 1975: Prestige speech styles: The imposed norm and inherent value hypothesis, in: Mc Cormack, W. C. and Wurm, S. (Ed.): *Language in Anthropology* 4, Language in many ways, Mouton [43]
- 224 Glenn, J. W. and Kleiner, N. 1968: Speaker identification based on nasal phonation, *J. acoust. soc. Amer.* 43, 368 - 372 [31 65 60]
- 225 Goerlitz, D. 1972: Ergebnisse und Probleme der ausdruckspsychologischen Sprechstimmforschung, Hain, Meisenheim [44]
- 226 Görttler, K. 1972: Stimme und Sprache, *Sitzungsber. d. Heidelberger Akad. d. Wiss.*, 189 - 195, Springer, Berlin [11 12]
- 227 Gon, J. J. D. v. d. and Hinte, N. v. 1959: The relation between the frequency of the alpha - rhythm and the speed of writing, *Electroenceph. clin. neurophysiol.* 11, 669 - 674 [12 66]
- 228 Goody, E. N. (Ed.) 1978: Questions and politeness: Strategies in social interaction, Cambridge univ. press [43 64]
- 229 Gottwald, W. 1982: Über Physiologie und Pathologie der Lautgebung (Phonation), *Sprache - Stimme - Gehör* 6, 118 - 124 [12 25]
- 230 Goydke, H. 1974: Über die Grundlagen und die Leistungsfähigkeit von Verfahren zur Sprecheridentifikation, *Taschenb. f. Kriminalisten* 24, Verlagsanst. Deutsche Polizei, Hilden [65]
- 231 Goydke, H. 1974: Untersuchungen über die Spektral - und Autokorrelationsanalyse von Sprache zum Zwecke der Sprecheridentifizierung, Diss., Fak. Maschinenbau u. Elektrotechnik, TU Braunschweig [52 60 65]
- 232 Gray, C. H. G. and Kopp, G. A. 1944: Voice print identification, unpubl. report, Bell laboratories [65]
- 233 Green, N. 1972: Automatic speaker recognition using pitch measurements in conversational speech, *Joint speech research unit report* 1000, Cheltenham, England [65]
- 234 Gregory, S. W. and Hoyt, B. R. 1982: Conversation partner mutual adaptation as demonstrated by Fourier series analysis, *J. psycholinguist. res.* 11, 35 - 46 [42 54]
- 235 Gross, M. 1980: Stimmfeldmessung, eine objektive Methode der Stimmdiagnostik, *Sprache - Stimme - Gehör* 4, 100 - 101 [25]
- 236 Grünewald, G. 1957: Bemerkungen zu einer vergleichenden Betrachtung von Sprechen und Schreiben, *Phonetica* 1, 193 - 202 [40 66]
- 237 Grünewald, G. 1957: Die Schreibdruck - Kurve, *Z. Menschenk.* 21, 133 - 177 [66]
- 238 Grünewald, G. 1957: Zur Schreib- und Sprechmotorik der Konstitutionstypen, *Z. Psychotherapie u. med. Psychol.* 7, 165 - 176 [19 54 66]

- 239 Grünewald, G. 1959: Über den Einfluß von Drogen auf die Schreibpsychomotorik, Arch. Psychiat. Z. Neurol. 198, 687 – 704 [22 66]
- 240 Grünewald, G., Zuberbier, E. and Spitznagel, A. 1957: Einführung zu den Vergleichsstudien zur Schreib- und Sprechmotorik, Z. Psychotherapie u. med. Psychol. 7, 162 – 164 [40 66]
- 241 Gubrynowicz, R. 1971: Methode d'analyse statistique du spectre de la parole: Application a la reconnaissance automatique du locuteur, Proc. 7th int. congr. Acoust. 3, 25 – 28 [65]
- 242 Gundermann, ?. 1982: Lernziel – Akupädie, Sprache – Stimme – Gehör 6, 79 – 81 [25 40]
- 243 Gupta, V. and Mermelstein, P. 1982: Effects of speaker on the performance of a speaker independent isolated – word – recognizer, J. acoust. soc. Amer. 71, 1581 – 1587 [32 34]
- 244 Habermann, G. 1978: Stimme und Sprache, Thieme, Stuttgart [64]
- 245 Hair, G. D. and Rekieta, T. W. 1972: Automatic speaker verification using phoneme spectra, J. acoust. soc. Amer. 51, 131 [31 65]
- 246 Hair, G. D. and Rekieta, T. W. 1973: Techniques for objective speaker identification, J. acoust. soc. Amer. 53, 354 [65]
- 247 Hall, C. S. and Lindzey, G. 1970: Theories of personality, Wiley & sons, New York [41]
- 248 Halle, M. et al. 1957: Acoustic properties of stop consonants, J. acoust. soc. Amer. 29, 107 – 116 [31]
- 249 Hanley, T. 1951: An analysis of vocal frequency and duration characteristics of selected samples from three American dialect regions, Speech monogr. 18, 78 – 93 [30 31]
- 250 Hanley, T. D., Snidecor, J. C. and Ringel, R. L. 1966: Some acoustic differences among languages, Phonetica 14, 97 – 107 [30]
- 251 Hardcastle, W. J. 1976: The physiology of speech production. An introduction for speech scientists, Acad. press, New York [12 64]
- 252 Hargreaves, W. A. 1961: Drug effects on vocal behavior, Amer. psychol. 16, 426 [22]
- 253 Hargreaves, W. A. and Starkweather, J. A. 1961: Vocal behavior: An illustrative case study, Meet. west. psychol. ass., Seattle [52 42]
- 254 Hargreaves, W. A. and Starkweather, J. A. 1963: Recognition of speaker identity, Lang. speech 6, 63 – 67 [65 52]
- 255 Hargreaves, W. A. and Starkweather, J. A. 1964: Voice quality changes in depression, Lang. speech 7, 84 – 88 [20 21 51]
- 256 Hargreaves, W. A., Starkweather, J. A. and Blacker, K. H. 1965: Voice quality in depression, J. abnormal soc. psychol. 70, 218 – 220 [21 52]
- 257 Harre, R. and Reynolds, V. (Ed) 1984: The meaning of primate signals, Cambridge University Press, Cambridge [9 31]
- 258 Harris, C. M. and Weiss, M. R. 1964: Effects of speaking condition on pitch, J. acoust. soc. Amer. 36, 933 – 936 [50]
- 259 Harris, F. J. 1978: On the use of windows for harmonic analysis with the discrete Fourier transform, Proc. of the inst. of electrical and electronics engineers 66, 51 – 83 [60]
- 260 Hart, R. J. and Brown, B. L. 1974: Personality information conveyed by the content and vocal aspects of speech, Speech monogr. 41, 371 – 380 [41]
- 261 Hartlieb, K. 1957: Der Umfang der Jugendstimme, Fol. phoniat. 9, 225 [15 50]
- 262 Hartlieb, K. 1962: Erbliche Merkmale der menschlichen Stimme, Z. menschl. Vererb. Konst. 36, 413 – 430 [13 50]
- 263 Hartlieb, K. 1969: Praktikum der Stimm- und Sprachheilhunde aus biokybernetischer Sicht, Reinhardt, München [25]
- 264 Hayre, H. S. 1983: Personality parameter detection from speech, Applied acoustics 16, 31 – 40 [41]
- 265 Hazen, B. 1973: Effects of differing contents on spectrographic speaker identification, J. acoust. soc. Amer. 54, 650 – 660 [51 65]

- 266 Hecker, M. H. L. 1971: Speaker recognition - an interpretative survey of the literature, Amer. speech and hearing association Monograph 16, Washington D. C. [63 65]
- 267 Heidrich, R. 1952: Experimentelle Untersuchungen über Körperbau, Stimme und Sprechweise, Z. menschl. Vererb. Konst. 31, 119-141 [19 55]
- 268 Heinemann, M. and Gabriel, H. 1982: Möglichkeiten und Grenzen der Stimmfeldmessung - Vorstellung des Heiserkeitsfeldes als Ergänzung der Methode, Sprache - Stimme - Gehör 6, 37-42 [20 25 50]
- 269 Helmholtz, H. v. 1913: Die Lehre von den Tonempfindungen, Vieweg, Braunschweig [55]
- 270 Henderson, J. B. et al. 1982: A cross-language study of velar coarticulation, J. acoust. soc. Amer. 71, S22 [12 31 30]
- 271 Heriot, A. 1956: The castrati in opera, Secker & Warburg, London [24]
- 272 Hermann, L. 1879: Weitere Untersuchungen über das Wesen der Vokale, Handbuch der Physiologie Bd. 1, Leipzig [31]
- 273 Herzog, H. 1933: Stimme und Persönlichkeit, Z. Psychol. 130, 300-379 [41]
- 274 Herzog, M. and Hopf, S. 1984: Behavioral responses to species-specific warning calls in infant squirrel monkeys reared in social isolation, Amer. J. of Primatology 7, 99-106 [9 42]
- 275 Hess, W. 1976: A pitch synchronous digital feature extraction system for phonemic recognition of speech, IEEE-ASSP 24, 14-25 [32]
- 276 Hess, W. 1983: Pitch determination of speech signals, Springer, Berlin [50]
- 277 Hesselmann, N. 1982: Computergestütztes visuelles Artikulationstraining, Sprache - Stimme - Gehör 6, 82-88 [25 51]
- 278 Hinde, R. A. 1974: Biological bases of human social behavior, Mc Graw-Hill, New York [42 43]
- 279 Hirano, M. 1974: Morphological structure of the vocal cord as a vibrator and its variations, Fol. phoniat. 26, 89-94 [10 11]
- 280 Hirano, M., Vennard, W. and Ohala, J. 1970: Regulation of register, pitch and intensity of voice: An electromyographic investigation of intrinsic laryngeal muscles, Fol. phoniat. 22, 1-20 [12 50 20]
- 281 Hoefker, U. et al. 1976: A new system for authentication of voice, Carnahan conference, Lexington, KY [65]
- 282 Hoefker, U. et al. 1981: Automatic recognition of speakers by computers, Pp. 100, Heinrich Hertz Inst. Nachrichtentechn., Berlin, PB82-147299 (7/82) [65]
- 283 Hoffmann, H. 1961: Stabilität und Faktorenstruktur einiger konventioneller Kriterien des Schreibdrucks, Psychol. Beitr. 6, 32-50 [66]
- 284 Hollien, H. 1960: Some laryngeal correlates of vocal pitch, J. speech hear. res. 3, 52-58 [11 50]
- 285 Hollien, H. 1960: Vocal pitch variation related to changes in vocal fold length, J. speech hear. res. 3, 150-156 [11 50]
- 286 Hollien, H. 1964: Laryngeal research by means of laminagraphy, Arch. Otolaryng. 80, 303 [11 50]
- 287 Hollien, H. 1974: On vocal registers, J. phonetics 2, 125-143 [20 51]
- 288 Hollien, H. and Hollien, P. A. 1971: A cross cultural study of adolescent voice change in European males, Proc. 7th int. congr. phonet. sci., Montreal, 332-337 [14 15]
- 289 Hollien, H. and Majewski, W. 1977: Speaker identification by long-term spectra under normal and distorted speech conditions, J. acoust. soc. Amer. 62, 975-980 [52 65]
- 290 Hollien, H. and Malcik, E. 1962: Adolescent voice change in southern negro males, Speech monogr. 29, 53-58 [15 17 50]
- 291 Hollien, H. and Malcik, E. 1967: Evaluation of cross-sectional studies of adolescent voice change in males, Speech monogr. 34, 80-84 [15 17 50]
- 292 Hollien, H. and McGlone, R. E. 1976: An evaluation of the voiceprint techniques for speaker identification, Proc. Carnahan conference, Lexington, KY [65]
- 293 Hollien, H. and Moore, G. P. 1960: Measurements of the vocal folds during changes in pitch, J. speech hear. res. 3, 157-165 [11 50]

- 294 Hollien, H. and Paul, P. 1969: A second evaluation of the speaking fundamental frequency characteristics of post-adolescent girls, *Lang. speech* 12, 119-124 [14 15 50]
- 295 Hollien, H. and Shipp, T. 1972: Speaking fundamental frequency and chronologic age in males, *J. speech hear. res.* 15, 155-159 [14 50]
- 296 Hollien, H., Malcik, E. and Hollien, B. 1965: Adolescent voice change in southern white males, *Speech monogr.* 32, 87-90 [15 17 50]
- 297 Holmgren, G. L. 1967: Physival and psychological correlates of speaker recognition, *J. speech hear. res.* 10, 57-66 [55 65]
- 298 Honjo, I. 1980: Laryngoscopic and voice characteristics of aged persons, *Arch. otolaryng.* 106, 149-150 [11 14 20]
- 299 Howes, W. L. 1979: Loudness of steady sounds - a new theory, *Acustica* 41, 277-320 [50]
- 300 Hudson, A. and Holbrook, A. 1982: Fundamental frequency characteristics of young black adults: Spontaneous speaking and oral reading, *J. speech hear. res.* 25, 25-28 [17 50]
- 301 Hunt, R. G. and Lin, T. K. 1967: Accuracy of judgments of personal attributes from speech, *J. personality soc. psychol.* 6 [41 55]
- 302 Husemann, A. J. 1982: *Der musikalische Bau des Menschen*, Verl. freies Geistesleben, Stuttgart [10 12]
- 303 Husson, R. 1960: *La voix chantée*, Gauthier-Villars, Paris [20]
- 304 Huttar, G. L. 1968: Relations between prosodic variables and emotions in normal American English utterances, *J speech hear. res.* 11, 481-487 [44 54 55]
- 305 Hymes, D. 1962: The ethnography of speaking, in: Gladwin, T. and Sturtevant, W. C. (Ed.): *Anthropology and human behavior*, Washington, D. C. [42 43]
- 306 Hymes, D. 1979: *Soziolinguistik: Zur Ethnographie der Kommunikation*, Suhrkamp, Frankfurt a. M. [43]
- 307 Imre, V. 1951: Veränderungen der Stimme während der Schwangerschaft, *Fol. phoniät.* 3, 224-232 [20 23 50]
- 308 Imre, V. 1968: Hormonell bedingte Stimmstörungen, *Fol. phoniät.* 20, 394-404 [20 25]
- 309 Ingemann, F. 1968: Identification of the speaker's sex from voiceless fricatives, *J. acoust. soc. Amer.* 44, 1142-1144 [16 31 55]
- 310 Ingrisano, D., Weismer, G. and Schuckers, G. H. 1980: Sex identification of preschool children's voices, *Fol. phoniät.* 32, 61-69 [15 16]
- 311 Irwin, R. B. 1977: Judgments of vocal quality, speech fluency and confidence of southern black and white speakers, *Lang. speech* 20, 261-266 [17 20]
- 312 Isshiki, N. 1964: Regulatory mechanism of voice intensity variation, *J. speech hear. res.* 7, 17-30 [12 50]
- 313 Jacobshagen, B. 1983: *Der Geschlechtsdimorphismus des menschlichen Stimmklangs im Spektralbild*, Vortrag, 18. Tg. Ges. Anthrop. Hum. Gen., Münster [16 52]
- 314 Jaschul, J. 1980: *Sprecheradaption mit optimierten Parametern für die automatische Spracherkennung*, DAGA 80, München, 727-730 [32 34]
- 315 Jaschul, J. 1982: *Adaption vorverarbeiteter Sprachsignale zum Erreichen der Sprecherunabhängigkeit automatischer Spracherkennungssysteme*, Diss, Fak. Elektrotechnik, TU München [32 34]
- 316 Jassem, W. 1971: Pitch and compass of the speaking voice, *J. int. phonet. Ass.* 1, 59-68 [34 50 54]
- 317 Jayant, N. S. 1974: Spectral analysis of speech based on linear delta modulation, in: Fant, G. (Ed): *Speech and Communication* 1, Almqvist & Wiksell, Stockholm, 129 [60]
- 318 Jones, H. E. 1942: The adolescent growth study 6: The analysis of voice records, *J. consul. psychol.* 6, 255-256 [15]
- 319 Jousselin, C. and Boe, L. J. 1982: Systeme adaptif de verification du locuteur, *Bull. inst. phonet. Grenoble* 10/11, 61-70 [60 65]
- 320 Jürgens, U. 1979: Vocalisation as an emotional indicator, *Behaviour* 69, 88-117 [9 44]

- 321 Kahane, J. C. 1983: Postnatal development and aging of the human larynx, *Seminars in speech and hearing* 4, 189–203 [11 14 15]
- 322 Kainz, F. 1961: Die "Sprache" der Tiere: Tatsachen, Problemschau, Theorie, Enke, Stuttgart [9]
- 323 Kallenbach, W. and Schröder, H. J. 1961: Zur Technik der Tonbandaufnahmen bei Sprechuntersuchungen, *Phonetica* 7, 95–108 [62]
- 324 Karlin, I. W. and Sobel, A. E. 1940: A comparative study of the blood chemistry of stutterers and non-stutterers, *Speech monogr.* 7, 75–84 [12 25 54]
- 325 Karolyi, L. v. 1971: *Anthropometrie*, Fischer, Stuttgart [19]
- 326 Kashyap, R. L. 1976: Speaker recognition from an unknown utterance and speaker speech interaction, *IEEE – ASSP* 24, 481–488 [65]
- 327 Kenneth, W. 1970: Talker recognition on large populations, *J. acoust. soc. Amer.* 47, 66 [51 65]
- 328 Kent, R. D. 1976: Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies, *J. speech hear. res.* 19, 421–447 [14]
- 329 Kent, R. D. 1978: Imitation of synthesized vowels by preschool children, *J. acoust. soc. Amer.* 63, 1193–1198 [15 31]
- 330 Kent, R. D. and Forner, L. L. 1979: Developmental study of vowel formant frequencies in an imitation task, *J. acoust. soc. Amer.* 65, 208–217 [31 15 51]
- 331 Kent, R. D. and Murray, A. D. 1982: Acoustic features of infant vocalic utterances at 3, 6 and 9 months, *J. acoust. soc. Amer.* 72, 353–365 [15 31]
- 332 Kersta, L. G. 1962: Voice print identification, *J. acoust. soc. Amer.* 34, 725 [51 65]
- 333 Kersta, L. G. 1962: Voiceprint – identification infallibility, *J. acoust. soc. Amer.* 34, 1978 [51 65]
- 334 Killian, W. 1984: Morphometrischer Sexualdimorphismus und Sexualdiagnose am tubulären Handskelett, *Z. f. Morphologie und Anthropologie* 75, 85–95 [16 19]
- 335 Kittel, G. and Hecht, L. 1977: Der erste Schrei – frequenzanalytische Untersuchungen, *Sprache – Stimme – Gehör* 1, 151–155 [15 25]
- 336 Kittel, G. and Moser, M. 1981: Objektivierung von Stimmunreinheiten mittels Computer, *Sprache – Stimme – Gehör* 5, 42–45 [20 50]
- 337 Kittel, G., Thürmer, S. and Thumfart, W. 1982: Das Kehlkopf – EMG als phoniatische Untersuchungsmethode für Prognose und Verlaufskontrolle von Stimm lippenlähmungen, *Sprache – Stimme – Gehör* 6, 114–117 [11 12]
- 338 Kiukaanniemi, H. and Mattila, P. 1980: Long-term speech spectra. A computerized method of measurement and a comparative study of Finnish and English data, *Scand. audiol.* 9, 67–72 [30 52]
- 339 Kiukaanniemi, H., Siponen, P. and Mattila, P. 1982: Individual differences in the long-term speech spectrum, *Fol. phoniatic.* 34, 21–28 [52]
- 340 Klingholz, F. and Martin, F. 1982: Die Erzeugung sehr niederfrequenter, quasiperiodischer Schwingungen im menschlichen Stimmapparat, *Sprache – Stimme – Gehör* 6, 13–15 [20 25]
- 341 Klingholz, F. and Martin, F. 1983a: Die quantitative Auswertung der Stimmfeldmessung, *Sprache – Stimme – Gehör* 7, 106–110 [50]
- 342 Klingholz, F. and Martin, F. 1983b: Zum Entstehungsmechanismus des vocal fry (VF), *Sprache – Stimme – Gehör* 7, 16–21 [11 20]
- 343 Kluge, E. and Steinwachs, F. 1952: Psychodiagnostische Studien an Schreib – und Griffdruck 2, *Z. Psychotherapie med. psychol.* 2, 230 [40 66]
- 344 Knußmann, R. 1961: Zur Methode der objektiven Körperbautypognoze, *Z. Konstitutionslehre* 36, 1 [19]
- 345 Knußmann, R. 1968: Entwicklung, Konstitution, Geschlecht, in: Becker, P. E. (Ed): *Humangenetik*, Bd. 1/1, 280–416, Thieme, Stuttgart [19]
- 346 Kohler, K. J., Schäfer, K. and Thou, W. 1981: Phonetische Untersuchungen zur gesprochenen Sprache im Deutschen, *Arbeitsber. Inst. Phonet. Univ. Kiel* 16, 109–136 [31]
- 347 Koida, G. 1980: Sprechererkennung mit Sprecheradaptationsparametern, *Dipl. – Arbeit, Lehrst. f. Datenverarb. TU München* [34 65]
- 348 Komiyama, S., Watanabe, H. and Ryu, S. 1984: Phonographic relationship between pitch and intensity of the human voice, *Fol. phoniatic.* 36, 1–7 [50]

- 349 Kramer, C. 1977: Perceptions of female and male speech, *Lang. speech* 20, 151–161 [16]
- 350 Kramer, E. 1963: Judgment of personal characteristics and emotions from nonverbal properties of speech, *Psychol. bull.* 60, 408–420 [41 44]
- 351 Kramer, E. 1964: Personality stereotypes in voice: A reconsideration of the data, *J. soc. Psych.* 62, 247–251 [41]
- 352 Krasner, M. et al. 1984: Investigation of text-independent speaker identification techniques under conditions of variable data, *ICASSP* 84, 18B.5.1 [65]
- 353 Kretschmer, E. 1956: *Medizinische Psychologie*, Thieme, Leipzig [19 21]
- 354 Kuenzel, H. J. and Borys, B. B. 1982: *Farbige Schallspektrographie: Einige Anwendungsbeispiele einer neuen Technik in der Phoniatrie, Sprache – Stimme – Gehör* 6, 74–78 [51]
- 355 Kuhn, M. H. 1981: On the word recognition potential of the long term spectrum, *Frequenz* 35, 190–194 [32 52]
- 356 Kuhn, M. H. 1982: Voice recognition, *Ewn. Electron.* 2(1), 52–59, *Ref. Zh. Fiz.* 11 Zh 940 (1982) [65]
- 357 Kunt, M. and Hugli, H. 1985: An overview of digital techniques for processing speech signals, in: *NATO ASI series, Vol. F16*, Mori, R. and Suen, C. Y. (Ed.): *New systems and architectures for automatic speech recognition and synthesis*, Springer, Heidelberg [32 60]
- 358 Kurtzberg, R. L., Alpert, M. and Friedhoff, A. J. 1963: Techniques for the reduction of trial-retrial variability, ??? [34 51 99]
- 359 La Riviere, C. 1974: Speaker identification from turbulent portions of fricatives, *Phonetica* 29, 246–252 [31 65]
- 360 Labov, W. 1963: The social motivation of a sound change, *Word* 19, 273–309 [43]
- 361 Labov, W. 1966: *The social stratification of English in New York city*, Center for applied linguistics, Washington [43]
- 362 Labov, W. 1976: *Sprache im sozialen Kontext*, Scriptor, Kronberg/Ts. [43]
- 363 Lacinä, O. 1968: Der Einfluß der Menstruation auf die Stimme der Sängerinnen – laryngopathia praemenstrualis, *Fol. phoniatri.* 20, 13–24 [11 23 20]
- 364 Ladefoged, P. 1978: Expectation affects identification by listening, *Lang. speech* 21, 373–374 [55 62 65]
- 365 Ladefoged, P. 1983: The limits of biological explanations in phonetics, in: Cohen, A. and Broecke, M. P. R. (Ed): *Abstracts of the 10th int. congr. phonet. sci.*, Foris Publ., Dordrecht, Holland, 31–37 [31 53]
- 366 Ladefoged, P. and Antonanzas, N. 1984: Computer measurements of breathy voice quality, *J. acoust. soc. Amer.* 75, S8 [20 60]
- 367 Ladefoged, P. and Bladon, A. 1982: Can a human speaker reproduce Fant's nomograms, *J. acoust. soc. Amer.* 71, S104 [53 31]
- 368 Lass, N. J. et al. 1976: Speaker sex identification from voiced, whispered and filtered isolated vowels, *J. acoust. soc. Amer.* 59, 675–678 [16]
- 369 Lass, N. J. et al. 1978: The effect of temporal speech alterations on speaker race and sex identifications, *Lang. speech* 21, 279–291 [16 17]
- 370 Lass, R. (Ed.) 1976: *Contemporary issues in experimental phonetics*, Academic press, New York [64]
- 371 Lastowka, M. and Sedlacek, K. 1984: *EMG – Untersuchungen der Kehlkopfmuskeln mittels einer Oberflächenelektrode*, *Sprache – Stimme – Gehör* 8, 8–10 [12 11 25]
- 372 Laver, J. 1979: *Voice quality: a classified bibliography*, Benjamins, Amsterdam [20 63]
- 373 Laver, J. 1980: *The phonetic description of voice quality*, Cambridge univ. press, Cambridge [20 63]
- 374 Lebrun, Y. 1983: *Cerebral dominance for language. A neurolinguistic approach*, *Fol. phoniatri.* 35, 13–39 [12 63]
- 375 Lejeune, F. E. 1964: Review of the available literature on the larynx for 1963, *Laryngoscope* 74, 1501–1534 and 1653–1675 [11 63]
- 376 Lenneberg, E. H. 1972: *Biologische Grundlagen der Sprache*, Suhrkamp, Frankfurt [9 10 12]

- 377 Lepley, W. M. 1952: The participation of implicit speech and acts of writing, *Amer. j. psychol.* 65, 597 – 600 [66]
- 378 Lersch, P. 1952: *Aufbau der Person*, Barth, München [40]
- 379 Li Chang – li, Song Zhi – yung and Zhao I – lung 1981: A new method for extracting fundamental frequencies from speech signals, *Chin. phys.* 1, 728 – 740 [50 60]
- 380 Li, K. P. and Hughes, G. W. 1973: Talker differences as they appear in correlation matrices derived from spectra of continuous speech, *J. acoust. soc. Amer.* 53, 355 [60 61 65]
- 381 Li, K. P. and Hughes, G. W. 1974: Talker differences as they appear in correlation matrices derived from spectra of continuous speech, *J. acoust. soc. Amer.* 55, 833 – 837 [60 61 65]
- 382 Li, K. P. and Wrench, E. H. 1982: Text – independent speaker recognition with short utterances, *J. acoust. soc. Amer.* 72, S29 – S30 [60 65]
- 383 Li, K. P., Dammann, J. E. and Chapman, W. D. 1966: Experimental studies in speaker verification using an adaptive system, *J. acoust. soc. Amer.* 40, 966 – 978 [65]
- 384 Lieberman, P. 1968: Primate vocalisations and human linguistic ability, *J. acoust. soc. Amer.* 44, 1574 – 1584 [9]
- 385 Lieberman, P. 1986: Some aspects of dimorphism and human speech, *Human Evolution* 1, 67 – 75 [15 16]
- 386 Lieberman, P. and Michaelis, S. B. 1962: Some aspects of fundamental frequency and envelope amplitude as related to the emotional content of speech, *J. acoust. soc. Amer.* 34, 922 – 927 [42 54]
- 387 Lieberman, P. et al. 1971: New born infant cry and nonhuman primate vocalisation, *J. speech hear. res.* 14, 718 – 727 [9 15]
- 388 Lieberman, P., Harris, K. S. and Wolff, P. 1968: Newborn infant cry in relation to nonhuman primate vocalisations, *J. acoust. soc. Amer.* 44, 365 [9 15]
- 389 Lieberman, P., Michaelis, S. B. and Soron, H. J. 1964: Emotional effects of fundamental frequency transformations, *J. acoust. soc. Amer.* 36, 1048 [42 67]
- 390 Lin, W. C. and Ganesan, K. 1976: Study of an on – line speaker independent word recognition system, based on acoustic – phonetic analysis and statistical pattern recognition techniques, *ICASSP 76*, Philadelphia, 194 – 197 [32]
- 391 Lindsley, C. F. 1934: The psycho – physical determinants of voice quality, *Speech monogr.* 1, 79 – 116 [12 20]
- 392 Linke, C. E. 1973: A study of pitch characteristics of female voices and their relationship to vocal effectiveness, *Fol. phoniat.* 25, 173 – 185 [16 50 20]
- 393 Lippert, K. L. and Biesalki, P. 1983: Klinische Erfahrungen mit einer neuen Technik der farbigen Schallspektrographie, *Sprache – Stimme – Gehör* 7, 112 [51]
- 394 Lowerre, B. T. 1977: Dynamic speaker adaptation in the Harpy speech recognition system, *ICASSP 77*, Hartford, 788 – 790 [34]
- 395 Lucci, V. 1978: Reconnaissance des variantes socioculturelles et situationnelles du français oral. Evaluation des parametres, *Bull. inst. phonet. Grenoble* 7, 33 – 66 [42 43]
- 396 Luchsinger, R. 1961: Die Sprachentwicklung von ein – und zweieiigen Zwillingen und die Vererbung von Sprachstörungen in den ersten drei Lebensjahren, *Fol. phoniat.* 13, 66 – 76 [15 18]
- 397 Luchsinger, R. and Arnold, O. E. 1959: *Lehrbuch der Stimm – und Sprachheilkunde*, Springer, Wien [64]
- 398 Luck, J. E. 1969: Automatic speaker verification using cepstral measurements, *J. acoust. soc. Amer.* 46, 1026 – 1032 [42 65]
- 399 Ludewig, M. 1971: Vergleich von Sprechstimmen mittels der Sonographie, *Diss med.*, München [51]
- 400 Lullies, H. 1953: Physiologie der Stimme und Sprache, in: Ranke, O. F. and Lullies, H.: *Lehrbuch der Physiologie. Gehör, Stimme und Sprache*, Springer, Heidelberg [64]
- 401 Lummis, R. C. 1973: Speaker verification by computer using intensity for temporal registration, *IEEE – AU 21*, 80 – 89 [65]

- 402 Lummis, R. C. and Rosenberg, A. E. 1971: Test of automatic speaker verification method with intensively trained professional mimics, *J. acoust. soc. Amer.* 51, 131 [42 31 65]
- 403 Mahnken, I. 1965: Zur Problematik der strukturellen Relevanz prosodischer Erscheinungen, in: Zwirner, E. (Ed.): *Proc. 5th int. congr. phonet. sci.*, 379–402, Karger, Basel [31 54]
- 404 Majewski, W. and Hollien, H. 1974: Euclidian distance between long-term speech spectra as a criterion for speaker identification, *Proc. speech communication seminar Stockholm* 3, 303–310 [52 61 65]
- 405 Majewski, W. and Hollien, H. 1975: Euclidean distance between long-term speech spectra, a criterion for speaker identification in: Fant, G.: *Speech and communication* 3, Almqvist & Wiksell, Stockholm, 303–310 [52 61 65]
- 406 Majewski, W., Hollien, H. and Zalewski, J. 1972: Speaking fundamental frequency of Polish adult males, *Phonetica* 25, 119–125 [19 50]
- 407 Makhoul, J. 1977: Linear prediction: A tutorial review, *Proc. IEEE* 63, 561–580 [60]
- 408 Mallory, E. and Miller, V. A. 1958: A possible basis for the association of voice characteristics and personality traits, *Speech monogr.* 25, 255–260 [41]
- 409 Malmberg, B. 1965: Stability and instability of syllabic structures, in: Zwirner, E. (Ed.): *Proc. 5th int. congr. phonet. sci.*, 403–408, Karger, Basel [31]
- 410 Mandelbaum, D. G. (Ed) 1949: *Selected writings of Edward Sapir*, Univ. of California Press, Berkeley [64]
- 411 Mangold, H. 1969: Die individuellen Unterschiede bei Langzeitmessungen am Sprachsignal, *Nachrichtentechnische Zeitschrift* 6, 364–367 [52]
- 412 Markel, J. D. and Davis, S. B. 1979: Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base, *IEEE-ASSP* 27, 74–82 [65]
- 413 Markel, J. D. and Davis, S. B. 1979: Text-independent speaker recognition from a large linguistically unconstrained time-spaced data base, *IEEE-ASSP* 27, 74–82 [65]
- 414 Markel, J. D. and Gray, A. H. 1976: *Linear prediction of speech*, Springer, Berlin [60]
- 415 Markel, J. D., Oshika, B. T. and Gray, A. H. 1977: Long term feature averaging for speaker recognition, *IEEE-ASSP* 25, 330–337 [65]
- 416 Markel, J. D., Oshika, B. T. and Gray, A. H. 1977: Long term feature averaging for speaker recognition, *IEEE-ASSP* 25, 330–337 [65]
- 417 Markel, N. N. et al. 1972: Personality traits associated with voice types, *J. psycholing. res.* 1, 249–255 [41]
- 418 Markel, N. N., Meisels, M. and Houck, J. E. 1964: Judging personality from voice quality, *J. abnormal soc. psychol.* 69, 458–463 [21 41]
- 419 Markel, N. N., Prebor, L. D. and Brandt, J. F. 1972: Biosocial factors in dyadic communication: sex and speaking intensity, *J. personality soc. psychol.* 23, 11–13 [16 42]
- 420 Marler, P. 1978: Vocal ethology of primates: Implications for psychophysics and psychophysiology, in: Chivers, D. J. and Herbert, J. (Ed): *Recent advances in primatology*, Vol. 1, Academic Press, London, 795–801 [9 12]
- 421 Martin, P. 1980: Variations prosodiques inter- et intra-locuteur, 11. *J. E. P. groupe communication parlee du GALF*, Strasbourg, 111–119 [54]
- 422 Martin, R. and Saller, K. 1957: *Lehrbuch der Anthropologie*, Bd. 1, Fischer, Stuttgart [10 19]
- 423 Martius, G., Steinwachs, F. and Jung, E. 1958: Untersuchungen über die distalmotorischen Bewegungsabläufe in der Schwangerschaft, *Z. menschl. Vererb. Konst.* 34, 533 [23 66]
- 424 Maskarinec, A. S. et al. 1981: Longitudinal observations of individual infant's vocalisations, *J. speech hear. dis.* 46, 267–273 [15]
- 425 Mathews, M. V., Miller, J. E. and David, E. E. 1961: Accurate estimate of the glottal waveshape, *J. acoust. soc. Amer.* 33, 843 [11 60]
- 426 Matsumoto, H. et al. 1973: Multidimensional representation of personal quality and its acoustical correlates, *IEEE-AU* 21, 428–436 [20]

- 427 Maurus, M. 1978: Problems arising and methods used in analysing the communicative meaning of the acoustic signals of squirrel monkeys, in: Chivers, D. J. and Herbert, J. (Ed): Recent advances in primatology, Vol. 1, Academic Press, London, 785-793 [9]
- 428 Mc Dermott, B. J. 1969: Multidimensional analysis of circuit quality judgments, J. acoust. soc. Amer. 45, 774-781 [55]
- 429 Mc Gehee, F. 1937: The reliability of the identification of the human voice, J. general psychol. 17, 249-271 [55 65]
- 430 Mc Gehee, F. 1944: An experimental study of voice recognition, J. general psychol. 31, 53-65 [55 65]
- 431 Mc Glone, R. E. 1966: Vocal pitch characteristics of children aged one to two years, Speech monogr. 33, 178-181 [15 50]
- 432 Mc Glone, R. E. and Hollien, H. 1963: Vocal pitch characteristics of aged women, J. speech hear. res. 6, 164-170 [14 50]
- 433 Mead, K. O. 1974: Identification of speakers from fundamental frequency contours in conversational speech, Joint speech research unit report 1002, Cheltenham, England [65]
- 434 Michael, W. and Crawford, C. C. 1927: An experiment in judging intelligence by the voice, J. educational psychol. 18, 107-114 [40]
- 435 Michel, J., Hollien, H. and Moore, P. 1966: Speaking fundamental frequency characteristics of 15, 16 and 17 year old girls, Lang. speech 9, 46-51 [15 50]
- 436 Miller, J. E. 1964: Decapitation and recapitation. A study of voice quality, J. acoust. soc. Amer. 36, 2002 [20 65]
- 437 Miller, M. H. 1963: Speech and voice patterns associated with hearing impairment, Maico audiological lib. ser. 2, 17-20 [20 62 67]
- 438 Milmoie, S. et al. 1967: The doctor's voice: Postdictor of successful referral of alcoholic patients, J. abnormal psychol. 72, 78-84 [41 42]
- 439 Milmoie, S., Novey, M. and Kagan, J. 1968: The mother's voice: Postdictor of aspects of her baby's behavior, Amer. psychol. ass., Proc. 76th ann. convention, 463-464 [15 42]
- 440 Mittleb, F. 1982: Voicing effect on vowel duration is not an absolute universal, J. acoust. soc. Amer. 71, S23 [30 31]
- 441 Monrad-Krohn, G. H. 1957: The third element of speech: Prosody in the neuropsychiatric clinic, J. ment. sci. 103, 326-331 [21 54]
- 442 Monsen, R. B. and Engebretson, A. M. 1977: Study of variations in the male and female glottal wave, J. acoust. soc. Amer. 62, 981-993 [11 62]
- 443 Monsen, R. B. and Engebretson, A. M. 1983: The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction, J. speech hear. res. 26, 89-97 [31 60]
- 444 Moran, M. J. and Gilbert, H. R. 1978: Speaking fundamental frequency characteristics of institutionalized adults with Down's syndrome, Amer. j. ment. deficiency 83, 248-252 [13 21 50]
- 445 Mori, R. de and Suen, C. Y. (Ed) 1985: New systems and architectures for automatic speech recognition and synthesis, Springer, Berlin [32 33 64]
- 446 Moser, M. 1984: Objektive Skalen der Stimmreinheit, Sprache - Stimme - Gehör 8, 35-37 [20]
- 447 Moses, P. J. 1941: Social adjustment and voice, Quart. j. speech 27, 532-536 [41 43]
- 448 Moses, P. J. 1941: Theories regarding the relation of constitution and character through the voice, Psychol. bull. 38, 747 [19 21 41]
- 449 Moses, P. J. 1942: The study of personality from records of the voice, J. consult. psychol. 6, 257-261 [41]
- 450 Moses, P. J. 1960: The psychology of the castrato voice, Fol. phoniat. 12, 204-216 [24 40]
- 451 Moskowitz, E. W. 1952: Voice quality in the schizophrenic reaction type, speech monogr. 19, 118-119 [17 21]
- 452 Mourant, A. E. and Morgan Watkin, I. 1952: Blood groups, anthropology and languages in Wales and the western countries, Heredity 6, 13 [13]

- 453 Müsebeck, K. and Rosenberg, H. 1983: Der phonatorische Luftstrom der Vokale, *Sprache - Stimme - Gehör* 7, 118 - 121 [12 62]
- 454 Muma, J. R., Leader, R. L. and Webb, C. E. 1968: Adolescent voice quality aberrations: personality and social status, *J. speech hear. res.* 11, 576 - 582 [20 43]
- 455 Mundinger, F. 1951: Zum Vererbungsproblem der menschlichen Singstimme, *Fol. phoniati.* 3, 191 - 223 [13 50]
- 456 Murry, T. and Doherty, E. T. 1980: Selected acoustic characteristics of pathologic and normal speakers, *J. speech hear. res.* 23, 361 - 369 [20 25]
- 457 Murry, T., Hoit-Dalgaard, J. H. and Gracco, V. L. 1983: Infant vocalisation: A longitudinal study of acoustic and temporal parameters, *Fol. phoniati.* 35, 245 - 253 [15]
- 458 Murry, T., Singh, S. and Sargent, M. 1977: Multidimensional classification of abnormal voice qualities, *J. acoust. soc. Amer.* 61, 1630 - 1635 [25 20 61]
- 459 Mysak, E. D. 1959: Pitch and duration characteristics of older males, *J. speech hear. res.* 2, 46 - 54 [14]
- 460 Nakagawa, S. et al. 1980: Considerations on speaker grouping by sex and age for automatic speech recognition, *Trans. Inst. Electron. Commun. Eng. Jpn.* E63 (12), 880, *Phys. Abstr.* 46636 (June 1981) [16 32]
- 461 Neiman, G. S., Klich, R. J. and Shuey, E. M. 1983: Voice onset time in young and 70-year-old women, *J. speech hear. res.* 26, 118 - 123 [14]
- 462 Ney, H. 1981: Measuring of the contours of intensity and fundamental period of speech for automatic speaker recognition, *Frequenz* 35 (10), 265 - 270 [54 65]
- 463 Ochiai, Y. 1959: Phoneme and voice identification studies using Japanese vowels, *Lang. speech* 2, 132 - 136 [31 65 62]
- 464 Ochiai, Y. and Fukumura, T. 1957: On the fundamental qualities of speech in communication, *J. acoust. soc. Amer.* 29, 392 - 393 [20 31]
- 465 Ohala, J. J. 1977: Speculations on pitch regulation, *Phonetica* 34, 310 - 312 [50]
- 466 Ohala, J. J. 1982: The voice of dominance, *J. acoust. soc. Amer.* 72, S66 [42]
- 467 Oordt, H. W. A. v. and Drost, H. A. 1963: Development of the frequency range of the voice in children, *Fol. phoniati.* 15, 289 - 298 [15 50]
- 468 Osgood, C. E., Suci, G. J. and Tannenbaum, P. H. 1957: The measurement of meaning, Univ. of Illinois Press, Urbana [40 43]
- 469 Ostwald, P. F. (Ed.) 1977: Communication and social interaction. Clinical and therapeutic aspects of human behavior, Grune & Stratton, New York [25 64]
- 470 Ostwald, P. F. 1963: Soundmaking: The acoustic communication of emotion, Thomas, Springfield [44]
- 471 Ostwald, P. F. 1965: Acoustic methods in psychiatry, *Sci. Amer.* 212 (3), 82 - 91 [22 21 52]
- 472 Ostwald, P. F. 1973: The semiotics of human sound, Mouton, The Hague [20 42 64]
- 473 Ostwald, P. F., Freeman, D. G. and Kurtz, J. H. 1962: Vocalisation of infant twins, *Fol. phoniati.* 14, 37 - 50 [15 18]
- 474 Pahn, J., Pahn, E. and Dettmann, R. 1980: Stimmtherapie mit Maskierungsgeräusch, *Sprache - Stimme - Gehör* 4, 150 - 153 [20 52]
- 475 Paliwal, K. K. 1982: On the performance of the frequency weighted cepstral coefficients in vowel recognition, *Speech communication* 1, 151 - 154 [32 60]
- 476 Paliwal, K. K. 1984: Effect of preemphasis on vowel recognition performance, *Speech communication* 3, 101 - 106 [31 60 61]
- 477 Paliwal, K. K. and Rao, P. V. S. 1982: Evaluation of various linear prediction parametric representations in vowel recognition, *Signal processing* 4, 323 - 327 [32]
- 478 Paul, J. E. 1976: Semi automatic speaker identification system (SASIS), final report. Rockwell international, Anaheim, Cal. [65]
- 479 Peters, R. W. 1954: Studies in extra messages: Listener identification of speaker's voices under conditions of certain restrictions imposed upon the voice signal, U. S. Naval school of aviation medicine joint project NM 001-064-01, Report 30, Pensacola, Florida [55 65]
- 480 Peterson, G. E. 1966: Glossary of terms from the physiological and acoustic phonetic theories, *J. speech hear. res.* 9, 100 - 120 []

- 481 Peterson, G. E. and Barney, H. L. 1952: Control methods used in a study of the vowels, *J. acoust. soc. Amer.* 24, 175 – 184 [31]
- 482 Pfau, W. 1973: Klassifizierung der menschlichen Stimme, Barth, Leipzig [20]
- 483 Phillis, J. A. 1970: Children's judgments of personality on the basis of voice quality, *Developmental psychol.* 3, 411 [41]
- 484 Ploog, D. 1974 : *ie Sprache der Affen und ihre Bedeutung für die Verständigungsweisen des Menschen*, Kindler, München [9]
- 485 Ploog, D. 1985: Stimme und Sprechen unter Kontrolle des Gehirns, in: Karlson, P. et al. (Ed.): *Information und Kommunikation, naturwissenschaftliche, medizinische und technische Aspekte*, Wiss. Verlagsgesellschaft., Stuttgart, 113 – 137 [9 12]
- 486 Pollack, I., Pickett, J. M. and Sumby, W. 1954: On the identification of speakers by voice, *J. acoust. soc. Amer.* 26, 403 [65 67]
- 487 Pollack, I., Rubenstein, H. and Horowitz, A. 1960: Communication of verbal modes of expression, *Lang. speech* 3, 121 – 130 [42 62 67]
- 488 Potter, R., Kopp, G. A. and Kopp, H. G. 1966: *Visible speech*, Dover publ., New York [51]
- 489 Pruzansky, S. 1963: Pattern – matching procedure for automatic talker recognition, *J. acoust. soc. Amer.* 35, 354 – 358 [52 65 61]
- 490 Pruzansky, S. and Mathews, M. V. 1964: Talker – recognition procedure based on analysis of variance, *J. acoust. soc. Amer.* 36, 2041 – 2047 [51 61 65]
- 491 Ptacek, P. H. and Sander, E. K. 1966: Age recognition from voice, *J. speech hear. res.* 9, 273 – 277 [14]
- 492 Ptacek, P. H. et al. 1966: Phonatory and related changes with advanced age, *J. speech hear. res.* 9, 353 – 360 [12 14]
- 493 Rabiner, L. R. 1978: On creating reference templates for speaker independent recognition of isolated words, *IEEE – ASSP* 26, 34 – 42 [32]
- 494 Rabiner, L. R. and Sambur, M. R. 1976: Speaker independent recognition of connected digits, *ICASSP* 76, Philadelphia, 202 – 205 [32]
- 495 Rabiner, L. R. and Schafer, R. W. 1978: *Digital processing of speech signals*, New Jersey [60]
- 496 Rabiner, L. R. and Wilpon, J. G. 1979: Application of clustering techniques to speaker trained isolated word recognition, *Bell system technical journal* 58, 2217 – 2233 [32 61]
- 497 Raes, J., Michelsson, K. and Kaiser, G. 1980: *Spektrographische Babyschreianalyse: Ein Hilfsmittel zur neuropädiatrischen Evaluation und Diagnose, Sprache – Stimme – Gehör* 4, 138 – 146 [15]
- 498 Raffler – Engel, W. v. 1965: Die Entwicklung vom Laut zum Phonem in der Kindersprache, in: *Proc. 5th int congr. phonet. sci.*, 482 – 485, Karger, Basel [15]
- 499 Ramig, L. A. and Ringel, R. L. 1983: Effects of physiological aging on selected acoustic characteristics of voice, *J. speech hear. res.* 26, 22 – 30 [14 20]
- 500 Reddy, D. R. 1976: Speech recognition by machine: A review, *Proc. IEEE* 46, 501 – 531 [32]
- 501 Regel, P. 1982: A module for acoustic – phonetic transcription of fluently spoken German speech, *IEEE – ASSP* 30, 440 – 450 [32]
- 502 Rekieta, T. W. and Hair, G. D. 1972: Mimic resistance of speaker verification using phoneme spectra, *J. acoust. soc. Amer.* 51, 131 [42 65]
- 503 Richter, H. 1962: Tradierung von Lautsystemen oder Vererbung? *Phonetica* 8, 73 – 92 [13 30]
- 504 Ringel, R. and Kluppel, D. 1964: Neo – natal crying. A normative study, *Fol. phoniat.* 16, 1 – 9 [15]
- 505 Roach, H. N. and Roach, P. J. 1983: Automatic recognition of languages, Joint speech research unit G. C. H. Q., Agreement Nr. P71/29/79, to appear in: *Work in progress*, Dpt. of Linguistics and Phonetics, Univ. Leeds [30 32]
- 506 Rosenberg, A. E. 1971: Effect of glottal pulse shape on the quality of natural vowels, *J. acoust. soc. Amer.* 49, 583 – 590 [11 31]
- 507 Rosenberg, A. E. 1973: Listener performance in speaker verification tasks, *IEEE – AU* 21, 221 – 225 [55 65]

- 508 Rosenberg, A. E. 1976: Automatic speaker verification, a review, Proc. IEEE 64/4, 475 – 487 [65]
- 509 Rosenberg, A. E. 1976: Automatic speaker verification: A review, Proc. IEEE 46, 475 – 487 [65]
- 510 Rosenberg, A. E. and Sambur, M. 1975: New techniques for automatic speaker verification, IEEE – ASSP 23, 169 – 176 [65]
- 511 Rosenberg, A. E. and Shipley, K. L. 1981: Speaker identification and verification combined with speaker independent word recognition, ICASSP 81, Atlanta, 184 – 187 [32 65]
- 512 Rostolland, D. 1982: Acoustic features of shouted voice, Acustica 50, 118 – 125 [20 50]
- 513 Rubenstein, L. 1961: Humming: a vocal standard with a diurnal variation, Science 134, 1519 – 1520 [42]
- 514 Rudert, I. 1965: Vom Ausdruck der Sprechstimme, in: Kirchhoff, R. (Ed): Ausdruckspsychologie, Bd. 5 Handbuch der Psychologie, Göttingen, 422 – 468 [40 44]
- 515 Rümman, F. 1977: Vergleichende sonographische Untersuchungen an Laienstimmen und ausgebildeten Stimmen, Med. dent. Diss., Univ. München [20]
- 516 Ruske, G. 1980: Automatische Erkennung gesprochener Sprache, DAGA 80, München, 61 – 72 [32]
- 517 Ruske, G. and Schotola, T. 1981: The efficiency of demisyllable segmentation in the recognition of spoken words, ICASSP 81, Atlanta, 971 – 974 [32]
- 518 Rutz, O. 1922: Sprache, Gesang und Körperhaltung, Beck, München [42 20]
- 519 Ryan, W. J. 1972: Acoustic aspects of the aging voice, J. Gerontology 27, 265 – 268 [14]
- 520 Ryan, W. J. and Burk, K. W. 1972: Predictors of age in the male voice, J. acoust. soc. Amer. 53, 345 [14]
- 521 Ryan, W. J. and Burk, K. W. 1974: Perceptual and acoustic correlates of aging in the speech of males, J. communication dis. 7, 181 – 192 [14]
- 522 Saal, R. 1977: Handbuch für das Funktionsmodell des Gehörs, Lehrst. f. Netzwerktheorie und Schaltungstechnik, TU München [55]
- 523 Sachs, L. 1978: Angewandte Statistik. Statistische Methoden und ihre Anwendung, Springer, Heidelberg [61]
- 524 Sachs, L. 1982: Statistische Methoden, Springer, Heidelberg [61]
- 525 Sakoe, H. and Chiba, S. 1978: Dynamic programming algorithm optimization for spoken word recognition, IEEE – ASSP 26, 43 – 49 [32]
- 526 Sambur, M. R. 1975: Selection of acoustic features for speaker identification, IEEE – ASSP 23, 176 – 182 [65]
- 527 Sambur, M. R. 1976: Speaker recognition using orthogonal linear prediction, IEEE – ASSP 24, 283 – 289 [65]
- 528 Sambur, M. R. 1976: Speaker recognition using orthogonal linear prediction, IEEE – ASSP 24, 283 – 289 [65]
- 529 Sanford, F. H. 1942: Speech and personality, Psychol. bull. 39, 811 – 845 [41]
- 530 Sapir, E. 1927: Speech as a personality trait, Amer. j. sociology 32, 892 – 905 [41]
- 531 Sato, S., Yokota, M. and Kasuya, H. 1982: Statistical relationships among the first three formant frequencies in vowel segments in continuous speech, Phonetica 39, 36 – 46 [31 51 61]
- 532 Savashima, M. and Cooper, F. S. (Ed.) 1977: Dynamic aspects of speech production, Univ. of Tokyo press, Tokyo [53 64]
- 533 Saxman, J. H. and Burk, K. W. 1967: Speaking fundamental frequency characteristics of middle – aged females, Fol. phoniat. 19, 167 – 172 [50]
- 534 Saxman, J. H. and Burk, K. W. 1968: Speaking fundamental frequency and rate characteristics of adult female schizophrenics, J. speech hear. res. 11, 194 – 203 [21]
- 535 Schaich, E. 1977: Schätz – und Testmethoden für Sozialwissenschaftler, München [61]
- 536 Scherer, K. R. (Ed.) 1982: Vokale Kommunikation. Nonverbale Aspekte des Sprachverhaltens, Beltz, Weinheim [42 63 64]

- 537 Scherer, K. R. 1970a: Attribution of personality from voice: A cross-cultural study on the dynamics of interpersonal perception, Unpubl. diss., Harvard univ., Dep. soc. relations, Cambridge, Mass. [41]
- 538 Scherer, K. R. 1970b: Non-verbale Kommunikation, Buske, Hamburg [42]
- 539 Scherer, K. R. 1972: Judging personality from voice: a cross-cultural approach to an old issue in interpersonal perception, *J. personality* 40, 191-210 [41]
- 540 Scherer, K. R. 1974a: Acoustic concomitants of emotional dimensions: Judging affects from synthesized tone sequences, in: Weitz, S. (Ed.): *Nonverbal communication*, New York, 105-111 [42]
- 541 Scherer, K. R. 1974b: Methoden der empirischen Sprachforschung, in: Koolwijk, J. van and Wicken-Mayser, M.: *Techniken der empirischen Sozialforschung* 3, Oldenbourg, München [43]
- 542 Scherer, K. R. 1978: Inference rules in personality attribution from voice quality: The loud voice of extraversion, *Europ. j. soc. psychol.* 8, 467-487 [41]
- 543 Scherer, K. R. and Giles, H. (Ed.) 1979: *Social markers in speech*, Cambridge univ. press, Cambridge [43 64]
- 544 Scherer, K. R. and Scherer, U. 1981: Speech behavior and personality, in: Darby, J. (Ed.): *Speech evaluation in psychiatry*, Grune & Stratton, New York, 115-135 [41]
- 545 Scherer, K. R., Koivumaki, J. and Rosenthal, R. 1972: Minimal clues in the vocal communication of affect: Judging emotions from context-masked speech, *J. psycholinguist. res.* 1, 269-285 [42]
- 546 Scherer, K. R., London, H. and Wolf, J. 1973: The voice of confidence. Paralinguistic cues and audience evaluation, *J. of research in personality* 7, 31-44 [41 55]
- 547 Scherer, K. R. 1974: Voice quality analysis of American and German speakers, *J. psycholing. res.* 3, 281-298 [20 41 43]
- 548 Scheuch, E. K. 1967: Auswahlverfahren in der Sozialforschung, Skalierungsverfahren in der Sozialforschung, in: König, R. (Ed.): *Handbuch der empirischen Sozialforschung*, Enke, Stuttgart, 309-384 [43 61]
- 549 Schick, A. 1979: *Schallwirkung aus physiologischer Sicht*, Klett-Cotta, Stuttgart [40 55]
- 550 Schilling, D. 1984: Song bouts and duetting in the Concolor Gibbon, in: Preuschoft, H. et al.: *The lesser apes*, Edinburgh University Press, Edinburgh [9 16]
- 551 Schilling, R. 1936: Über die Stimme erbgleicher Zwillinge, *Klin. Wochenschr.* 15, 756-757 [18]
- 552 Schilling, R. 1950: Über die Stimme erbgleicher Zwillinge, *Fol. phoniat.* 2, 98-119 and 205-238 [18]
- 553 Schilling, R. 1952: Umwandlung unserer stimmphysiologischen Vorstellungen aufgrund der Görttler'schen anatomischen Untersuchungen an menschlichen Stimmband, *Z. Phonetik* 6, 94-105 [11 12]
- 554 Schneewind, K. A. 1966: Über den Zusammenhang von Selbst- und Fremdeinschätzung bei der Beurteilung menschlicher Sprechstimmen, *Diss. Phil.*, Erlangen-Nürnberg [40 55]
- 555 Schotola, T. 1980: Silbenanlautende und silbenauslautende Konsonantenfolgen als Entscheidungseinheiten für die automatische Spracherkennung, *Diss TU München* [32]
- 556 Schröder, M. R. 1967: Determination of the geometry of the human vocal tract by acoustic measurements, *J. acoust. soc. Amer.* 41, 1002-1010 [10 53]
- 557 Schröder, M. (Ed.) 1985: *Speech and speaker recognition*, Karger, Basel [32 65]
- 558 Schubö, W. and Uehlinger, H. M. 1984: *SPSSx, Handbuch der Programmversion 2*, Fischer, Stuttgart [61]
- 559 Schuessler, W. (Ed) 1973: *Signalverarbeitung*, Erlangen [60]
- 560 Schultz-Coulon, H. J. 1975: Grundtonanalyse - ein Beitrag zur objektiven Beurteilung der Sprech- und Singstimme, *HNO* 23, 218 [20 50]
- 561 Schwarz, M. F. 1968: Identification of speaker sex from isolated voiceless fricatives, *J. acoust. soc. Amer.* 43, 1178 [16 31]

- 562 Schwartz, M. F. and Rine, H. E. 1968: Identification of speaker sex from isolated, whispered vowels, *J. acoust. soc. Amer.* 44, 1736 - 1737 [16 31]
- 563 Schweinsberg, F. 1946: Stimmliche Ausdrucksgestaltung im Dienste der Kirche, Heidelberg [43]
- 564 Segre, R. 1971: Senescence of the voice, *Eye ear nose and throat monthly* 50, 223 - 227 [14 50]
- 565 Seidner, W., Stürzebecher, E. and Wagner, H. 1981: Belastungsphonographie, *Fol. phoniat.* 33, 100 - 104 [25]
- 566 Seip, H. S. 1946: A study in listener reaction to voice quality, *Speech monogr.* 13, 120 - 121 [41 20 67]
- 567 Seyfarth, R. M. 1984: What the vocalisations of monkeys mean to humans and what they mean to the monkeys themselves, in: Harre, R. and Reynolds, V. (Ed): *The meaning of primate signals*, Cambridge University Press, Cambridge [9]
- 568 Seymour, H. N. and Seymour, C. M. 1981: Black English and standard American English in consonantal development of four- and five-year-old children, *J. speech hear. dis.* 46, 274 - 280 [17]
- 569 Shearme, J. N. and Holmes, J. N. 1959: An experiment concerning the recognition of voices, *Lang. speech* 2, 123 - 131 [65 67]
- 570 Shipp, T. and Hollien, H. 1969: Perception of the aging male voice, *J. speech hear. res.* 12, 703 - 710 [14 67]
- 571 Shrinidhar, M. and Mohankrishnan, N. 1982: Text-independent speaker recognition: A review and some new results, *Speech communication* 1, 257 - 267 [65]
- 572 Simonov, P. V. and Frolov, M. V. 1973: Utilization of human voice for estimation of man's emotional stress and state of attention, *Aerospace medicine* 44, 265 - 258 [44]
- 573 Singh, S., Woods, D. and Tishman, A. 1972: An alternative MD-scale analysis of the Graham and House data, *J. acoust. soc. Amer.* 51, 666 - 668 [15 61 67]
- 574 Sirviö, P. and Michelsson, K. 1976: Sound-spectrographic cry analysis of normal and abnormal newborn infants. A review and a recommendation of the cry characteristics, *Fol. phoniat.* 28, 161 - 173 [15 25 63]
- 575 Skinner, E. R. 1935: A calibrated recording and analysis of the pitch, force and quality of vocal tones expressing happiness and sadness, *Speech monogr.* 2, 81 - 137 [12 42]
- 576 Smith, B. L., Brown, B. L., Strong, W. J. and Rencher, A. C. 1975: Effects of speech rate on personality perception, *Lang. speech* 18, 145 - 152 [41 54]
- 577 Smith, J. E. Keith 1962: Decision theoretic speaker recognizer, *J. acoust. soc. Amer.* 34, 1988 [61 65]
- 578 Smith, J. E. Keith and Klem, L. 1961: Vowel recognition using a multiple discriminant function, *J. acoust. soc. Amer.* 33, 358 [32 61]
- 579 Smrkowski, L. 1977: Voice identification, Michigan Dep. of state police [65]
- 580 Smrkowski, R. 1976: Study of speaker identification by aural and visual examination of non-contemporary speech samples, *J. assoc. official analytical chemists* 59, 927 - 931 [51 65]
- 581 Sneath, P. H. A. and Sokal, R. 1973: Numerical taxonomy: The principles and practice of numerical classification, San Francisco [61]
- 582 Snidecor, J. C. 1943: A comparative study of pitch and duration characteristics of impromptu speaking and oral reading, *Speech monogr.* 10, 50 - 56 [50 54]
- 583 Snowdon, C. T. and Pola, Y. V. 1978: Responses of pygmy Marmosets to synthesized variations of their own vocalisations, in: Chivers, D. J. and Herbert, J. (Ed): *Recent advances in primatology*, Vol. 1, Academic Press, London, 811 - 813 [9]
- 584 Sodeur, W. 1974: Empirische Verfahren zur Klassifikation, Stuttgart [61]
- 585 Sondhi, M. M. 1975: Measurement of the glottal waveform, *J. acoust. soc. Amer.* 57, 228 - 232 [11 62]
- 586 Soskin, W. and Kauffman, P. 1961: Judgment of emotion in word-free voice samples, *J. communication* 11, 73 - 80 [42]

- 587 Sovijärvi, A. 1965: Die Bestimmung der Stimmkategorien mittels Resonanzröhren, in: Zwirner, E. (Ed.): Proc. 5th int. congr. phonet. sci., Karger, Basel, 532–535 [25 50]
- 588 Späth, Helmut 1975: Cluster – Analyse – Algorithmen zur Objektklassifizierung und Datenreduktion, München [61]
- 589 Spillmann, T., Dillier, N. and Güntensperger, J. 1979: Zur Modifikation von Stimmparametern durch ein linear – prediction – Modell: Eine Methode der digitalen Analyse und Synthese von Sprache, Sprache – Stimme – Gehör 3, 46–54 [33 53 60]
- 590 Spirduso, W. W. 1980: Physical fitness: aging and psychomotor speed: A review, J. Gerontology 35, 850–865 [14]
- 591 Spitznagel, A. 1959: Zur Schreib – und Sprechmotorik der Epileptiker, Z. Psychotherapie und med. Psychologie 8, 30–44 [21 42 66]
- 592 Starkweather, J. A. 1956: Content – free speech as a source of information about the speaker, J. abnormal psychol. 52, 394–402 [12 42]
- 593 Starkweather, J. A. 1956: The communication value of content – free speech, Amer. j. psychol. 69, 121–123 [42]
- 594 Starkweather, J. A. 1961: Vocal communication of personality and human feelings, J. communication 11, 63–72 [41 42]
- 595 Starkweather, J. A. and Hargreaves, W. A. 1964: The influence of sodium pentobarbital on vocal behavior, J. abnormal and soc. psychol. 69, 123–126 [22 52]
- 596 Stebbins, W. C. 1978: Hearing and acoustic communication, Introduction, in: Chivers, D. J. and Herbert, J. (Ed): Recent advances in primatology, Vol. 1, Academic Press, London, 703–704 [9]
- 597 Steffen – Batog, M., Jassem, W. and Gruszka – Koscielak, H. 1970: Statistical distributions of short – term F – values as a personal voice characteristic, in: Jassem, W.: Speech analysis and synthesis, Vol. 2, Warsaw, 197–208 [20 50]
- 598 Steinwachs, F. 1950: Bericht aus dem psychologischen Laboratorium der Univ. Nervenlinik Tübingen über den psychomotorischen Schreibtest, Psychol. Rundschau 4, 250–251 [66]
- 599 Steinwachs, F. 1952: Psychodiagnostische Studien an Schreib – und Griffdruck 1: Konstitutionstypische Grundkurven der Handschrift und ihre pathologischen Veränderungen, Z. Psychotherapie und med. Psychologie 2, 41–47 [19 66]
- 600 Steinwachs, F. 1954: Schreibmotorik und Schreibmaterial bei Grundschulkindern. Grundlagen zur Psychomotorik der Handschrift, Hogrefe, Göttingen [66]
- 601 Steinwachs, F. 1962: Exakte feinmotorische Diagnostik der individuellen Temperamentskonstanten mit ihren exogenen Variabilitäten, Z. menschl. Vererb. Konst. 36, 191–204 [66]
- 602 Steinwachs, F. and Barmeyer, H. 1952: Die Beziehungen der Feinmotorik zu den puberalen Alters – und Reifungsgraden, Z. menschl. Vererb. Konst. 31, 174 [66]
- 603 Steinwachs, F. and Kluge, E. 1952: Psychodiagnostische Studien an Schreib – und Griffdruck 2: Die sphärisch – bildhafte Beeinflussung der Handschrift in Hypnose, Z. Psychotherapie und med. Psychologie 2, 230–234 [66]
- 604 Steller, W. 1959: Generationsunterschiede im Friesischen, Phonetica 3, 39–50 [30 31]
- 605 Sternberg, S. 1969: Memory scanning: Mental processes revealed by reaction – time experiments, Amer. scientist 57, 421–457 [55]
- 606 Stevens, K. H. et al. 1968: Speaker authentication and identification: A comparison of spectrographic and auditory presentation of speech materials, J. acoust. soc. Amer. 44, 1596–1607 [65]
- 607 Stevens, K. N. 1971: Source of inter – and intra – speaker variability in the acoustic properties of speech sounds, Proc. 7th int. congr. phonet. sci., Montreal, 206–232 [31 65]
- 608 Stevens, K. N. 1972: Sources of inter and intra speaker variability in the acoustic properties of speech sounds, Proc. 7th int. congr. of phonetic sciences, A. Rigault and R. Charbonneau (Ed), Mouton, The Hague, 206–232 [65]
- 609 Stevens, K. N. 1980: Acoustic correlates of some phonetic categories, J. acoust. soc. Amer. 68, 836–842 [31]

- 510 Stoicheff, M. L. 1981: Speaking fundamental frequency characteristics of nonsmoking female adults, *J. speech hear. res.* 24, 437–441 [50]
- 511 Strong, W. J. 1967: Machine aided formant determination for speech synthesis, *J. acoust. soc. Amer.* 41, 1434–1442 [33]
- 512 Strong, W. J. and Purves, R. B. 1972: Scheme for automatic formant analysis of speech, *J. acoust. soc. Amer.* 51, 110 [32]
- 513 Studdert-Kennedy, M. et al. 1981: Hemispheric specialisation for language processes, *Science* 211(4485), 960–961 [12]
- 514 Stumpf, C. 1926: *Die Sprachlaute*, Berlin [31]
- 515 Su, L.-S., Li, K.-P. and Fu, K. S. 1974: Identification of speakers by use of nasal coarticulation, *J. acoust. soc. Amer.* 56, 1876–1882 [65]
- 516 Sundberg, J. 1973: The source spectrum in professional singing, *FoL phoniat.* 25, 71–90 [11 20]
- 517 Sychra, A. and Sedalacek, K. 1967: Relations between the acoustic, articulatory and psychological parameters of the emotional expressions in speech, *Proc. 6th int. Congr. phonet. sci.*, Prague, 811–812 [44]
- 518 Tarnoczy, T. 1961: Über das individuelle Sprachspektrum, *Proc. 4th int. Congr. of phonetic sciences*, Helsinki, 259–264 [31 51]
- 519 Tarnoczy, T. H. 1947: Physical characteristics of speech sounds and some aspects of their anthropological relations, *Acta anthropologica*, Budapest, 1, 1–43 [31 10]
- 520 Tatsuoka, M. M. 1971: *Multivariate analysis. Techniques for educational and psychological research*, Wiley & sons, New York [61]
- 521 Taylor, H. C. 1934: Social agreements on personality traits as judged from speech, *J. soc. psychol.* 5, 244–248 [41]
- 522 Terango, L. 1966: Pitch and duration characteristics of the oral reading of males on a masculinity–femininity dimension, *J. speech hear. res.* 9, 590–595 [16 50]
- 523 Thompson, R. W. 1958: The "th"–sounds and genetics, *Phonetica* 2, 184–192 [13 30]
- 524 Tillmann, H. G. 1967: Automatische Identifikation von Sprechern, *Nachrichten-technische Zeitschrift* 12, 706–713 [65]
- 525 Tillmann, H. G. 1974: Das individuelle Subjekt und seine Identität im phonetischen Kommunikationsprozeß, Buske, Hamburg [31 65]
- 526 Tillmann, H. G. and Mansell, P. 1980: *Phonetik. Lautsprachliche Zeichen, Sprachsignale und lautsprachlicher Kommunikationsprozeß*, Klett, Stuttgart [31 64]
- 527 Tischler, H. 1957: Schreien, Lallen und erstes Sprechen in der Entwicklung des Säuglings, *Z. Psychologie* 160, 209–263 [15]
- 528 Titze, I. R. 1973: The human vocal cords: a mathematical model, Part 1, *Phonetica* 28, 129–170 [11 53]
- 529 Titze, I. R. 1974: The human vocal cords: a mathematical model, Part 2, *Phonetica* 29, 1–21 [11 53 63]
- 530 Titze, I. R. 1976: Stresses in vocal fold tissues during normal phonation, *J. acoust. soc. Amer.* 60, 564 (A) [11 12]
- 531 Titze, I. R. and Strong, W. J. 1975: Normal modes in vocal cord tissues, *J. acoust. soc. Amer.* 57, 736–744 [11 12]
- 532 Titze, I. R. and Talkin, D. T. 1979: A theoretical study of the effect of various laryngeal configurations on the acoustics of phonation, *J. acoust. soc. Amer.* 66, 60–74 [11 53]
- 533 Todt, E. and Howell, R. J. 1980: Vocal cues as indices of schizophrenia, *J. speech hear. res.* 23, 517–526 [21]
- 534 Toledano, S. 1977: Smallest space analysis, a new subprogram "SSA1" in the SPSS–package, Computation Center, Hebrew university of Jerusalem [61]
- 535 Tosi, O. 1979: *Voice identification. Theory and legal applications*, University Park Press, Baltimore [65]
- 536 Tosi, O. and Postan, D. 1974: Longitudinal study of glottal frequencies of children at the age of puberty, *FoL phoniat.* 26, 227–228 [15 50]
- 537 Tosi, O. et al. 1972: Experiment on voice identification, *J. acoust. soc. Amer.* 51, 2030–2043 [65 51]

- 633 Toussaint, G. T. 1974: Bibliography on estimation of misclassification, IEEE -IT 20, 472-479 [63 61]
- 639 Trager, G. L. 1961: The typology of paralinguage, Anthropological linguistics 3, 17-21 [42 20]
- 640 Traunmüller, H. 1984: Articulatory and perceptual factors controlling the age- and sex-conditioned variability in formant frequencies of vowels, Speech communication 3, 49-61 [14 16 31]
- 641 Trendelenburg, F. 1950: Einführung in die Akustik, Berlin [53]
- 642 Trojan, F. 1952: Der Ausdruck der Sprechstimme, Maudrich, Wien [42]
- 643 Trojan, F. 1952: Experimentelle Untersuchungen über den Zusammenhang zwischen dem Ausdruck der Sprechstimme und dem vegetativen Nervensystem, Fol. phoniat. 4, 65-92 [12 42]
- 644 Trojan, F. 1954: Der Ausdruck der Sprechstimme im Deutschen, Diesterweg, Frankfurt [42 30]
- 645 Trojan, F. 1959: Die Ausdruckstheorie der Sprechstimme, Phonetica 4, 121-150 [42 63]
- 646 Trojan, F. 1960: Die Stimme des Hypnotiseurs, Fol. phoniat. 12, 137-144 [40 42]
- 647 Trojan, F. 1965: Grundlagen der Biophonetik, in: Zwirner, E. (Ed.): Proc. 5th congr. phonet. sci., Karger, Basel, 544-547 [42]
- 648 Trojan, F. 1975: Biophonetik, Bibliogr. Inst. Mannheim [42 12]
- 649 Troland, L. T. 1929: The psychophysiology of auditory qualities and attributes, J. general psychol. 2, 28-58 [67]
- 650 Trudgill, P. 1974: The social differentiation of English in Norwich, Cambridge University Press, Cambridge [30 43]
- 651 Ungeheuer, G. 1977: Sprache und Signal, Buske, Hamburg [31 60]
- 652 Vemula, N. R. et al. 1979: A speech microscope, J. acoust. soc. Amer. 65, S22 [51 60]
- 653 Vernon, C. et al. 1977: An objective method for the spectrographic analysis of sounds, Lang. speech 20, 67-72 [51]
- 654 Verschuure, J. and Meeteren, A. A. van 1975: The effect of intensity on pitch, Acustica 32, 33-44 [50]
- 655 Voiers, W. D. 1964: Perceptual bases of speaker identity, J. acoust. soc. Amer. 36, 1065-1073 [65 67]
- 656 Voiers, W. D. 1965: Performance evaluation of speech processing devices 2, the role of individual differences, Report AFCRL-66-24, Air force Cambridge research laboratories, Office of aerospace research, Bedford, Mass., Proc. IEEE 1964, 460-475 [32]
- 657 Voiers, W. D., Cohen, M. F. and Mickunas, J. 1965: Evaluation of speech processing devices 1: Intelligibility, quality, speaker recognizability, Report AFCRL-65-826, Air force Cambridge research laboratories, Office of aerospace research, Bedford, Mass. [32 65]
- 658 Wängler, H. H. 1968: Atlas deutscher Sprachlaute, Akademie Verlag, Berlin [31]
- 659 Wallner, L. J. et al. 1968: Voice changes following adenotonsillectomy, Laryngoscope 78, 1410-1418 [25]
- 660 Wasz-Höckert, O. et al. 1968: Infant cry. A spectrographic and auditory analysis, Heinemann, London [15]
- 661 Wayne, M. C. 1978: The relationship between measures of nasality and measures of the oral and nasal sound pressure ratio, Lang. speech 21, 69-75 [12 20]
- 662 Wedin, S. and Ogren, J.-E. 1982: Analysis of the fundamental frequency of the human voice and its frequency distribution before and after a voice training program, Fol. phoniat. 34, 143-149 [20 52 50]
- 663 Wedin, S., Leanderson, R. and Wedin, L. 1978: Evaluation of voice training - spectral analysis compared with listener's judgments, Folia phoniatrica 30, 103-112 [20 55]
- 664 Weihs, H. 1958: Psychologie der Stimmstörungen, Z. Menschenkunde 22, 189 [40 25]
- 665 Weinberg, B. and Bennet, S. 1971: A study of talker sex recognition of esophageal voices, J. speech hear. res. 14, 391-395 [16 25]

- 666 Weinberg, B. and Bennet, S. 1971: Speaker sex recognition of 5- and 6-year-old children's voices, *J. acoust. soc. Amer.* 50, 1210-1213 [15 16]
- 667 Weinberg, B. and Zlatin, M. 1970: Speaking fundamental frequency characteristics of mongoloid girls, *J. speech hear. res.* 13, 418-425 [25 50]
- 668 Weiske, T. 1943: Sprechdruck und Erbcharakter, *Z. angew. Psychol. u. Charakterkunde* 65, 51-97 and 113-168 [40 42 13]
- 669 Weiss, D. A. 1950: The pubertal change of the human voice, *Fol. phoniatic.* 2, 126-159 [15 63]
- 670 Weitz, S. 1974: Nonverbal communication. Readings with commentary, Oxford Univ. press, New York [42 64]
- 671 Welch, P. D. and Wimpess, G. E. 1961: Two multivariate statistical computer programs and their application to the vowel recognition problem, *J. acoust. soc. Amer.* 33, 426 [32 61]
- 672 Wendahl, R. W. 1963: Laryngeal analog synthesis of harsh voice quality, *Fol. phoniatic.* 15, 241-250 [20]
- 673 Wendahl, R. W., Moore, P. and Hollien, H. 1963: Comments on vocal fry, *Fol. phoniatic.* 15, 251-255 [20]
- 674 Werner, H. 1932: Grundlagen der Sprachphysiognomik, Barth, Leipzig [12 42]
- 675 Whitaker, H. H. 1976: Neurobiology of language, in: Carterette, E. C. and Friedman, M. P. (Ed.): Handbook of perception, Vol. 3: Language and speech, Academic Press, New York, 121-144 [12 42]
- 676 Whitehead, R. L. and Emanuel, F. W. 1974: Some spectrographic and perceptual features of vocal fry, abnormally rough and modal register vowel phonations, *J. communication dis.* 7, 305-319 [20 51]
- 677 Wieser, M. and Schlorhafer, W. 1977: Objektive Meßmethoden bei Stimmstörungen am Beispiel der Mutationsfistelstimmen, *Sprache - Stimme - Gehör* 1, 127-134 [25 60 51]
- 678 Wieser, M. and Schlorhafer, W. 1978: Spektrographische Untersuchungen bei spastischen Dysphonien, *Sprache - Stimme - Gehör* 2, 86-95 [25]
- 679 Wilcox, K. A. and Horii, Y. 1980: Age and change in vocal jitter, *J. Gerontology* 35, 194-198 [14 20]
- 680 Wilkens, H.H. and Bartel, H. H. 1977: Wiedererkennbarkeit der Originallautstärke eines Sprechers bei elektroakustischer Wiedergabe, *Acustica* 37, 45-49 [50]
- 681 Williams, C. E. 1964: The effects of selected factors on the aural identification of speakers, Sect. 3 of report EST-TDR-65-153, Electronics systems division, Air force system command, Hanscom field [55 65]
- 682 Williams, C. E. and Stevens, K. N. 1970: Acoustical manifestation of emotional speech, *J. acoust. soc. Amer.* 47, 65 [42]
- 683 Williams, C. E. and Stevens, K. N. 1972: Emotions and speech: Some acoustical correlates, *J. acoust. soc. Amer.* 52, 1238-1250 [42 51]
- 684 Williamson, D. G. and Harmon-Smith, A. 1980: Der Einfluß der Grundfrequenz auf die Sprachverständlichkeit, *Audiol. Akust.* 19, 236-240 [50 31]
- 685 Winckel, F. and Krause, M. 1965: Ermittlung von Spracheigenschaften aus statistischen Verteilungen von Amplitude und Tonhöhe, *Proc. 5th int. congr. acoust., Liege, Paper A41* [31 50]
- 686 Winner, A. - R. 1983: Zusammenhänge zwischen Intensität, Tonhöhe und spektraler Energieverteilung bei individuellen Sprechern - untersucht am Beispiel isolierter Vokale ausgewählter Sprecher, unpubl. Magisterarbeit, Inst. Phonetik u. sprachl. Kommunikation, Univ. München [20]
- 687 Wolf, J. J. 1972: Efficient acoustic parameters for speaker recognition, *J. acoust. soc. Amer.* 51, 2044-2056 [65]
- 688 Wrench, E. H. 1980: Speaker authentication operational test and evaluation, ITT defense communications division RADC - TR - 80 - 64, final technical report [65]
- 689 Wyke, B. (Ed) 1974: Ventilatory and Phonatory control mechanisms, Oxford University Press, London [12]
- 690 Yasuhiro, T. 1982: Experiment on male-to-female voice conversion, *Acustica* 51, 273-279 [16 20]

- 691 Young, M. A. and Campbell, R. A. 1967: Effects of context on talker identification, J. acoust. soc. Amer. 42, 1250 – 1254 [65]
- 692 Yumoto, E. 1983: The quantitative evaluation of hoarseness, a new harmonics to noise ratio method, Arch. Otolaryngology 109, 48 – 52 [20]
- 693 Yumoto, E., Gould, W. J. and Baer, T. 1982: Harmonics – to – noise ratio as an index of the degree of hoarseness, J. acoust. soc. Amer. 71, 1544 – 1550 [20]
- 694 Yumoto, E., Sasaki, Y. and Okamura, H. 1984: Harmonics – to – noise ratio and psychophysical measurement of the degree of hoarseness, J. speech hear. res. 27, 2 – 6 [20]
- 695 Zalewski, J., Majewski, W. and Hollien, H. 1975: Cross – correlation of long – term speech spectra as a speaker identification technique, Acoustica 34, 20 – 24 [52 65 61]
- 696 Zenker, W. and Zenker, A. 1960: Über die Regelung der Stimmlippenspannung durch von außen eingreifende Mechanismen, Fol. phoniat. 12, 1 – 36 [11 12]
- 697 Zerling, J. P. 1980: Correlation entre variabilité articulaire et acoustique chez deux locuteurs, 11. J. E. P. groupe communication parlée du GALF, Strasbourg, 121 – 132 [31 54]
- 698 Zuberbier, E. 1957: Zur Schreib – und Sprechmotorik der Depressiven, Z. Psychotherapie med. Psychol. 7, ?? [21 42 66 99]
- 699 Zuberbier, E. 1960: Über die Einwirkung von Stimmungsfaktoren auf Sprech – und Schreibweise, Psychol. Beitr. 4, 253 – 272 [42 54 66]
- 700 Zuberbier, E. and Grünewald, G. 1960: Studien zur Psychomotorik der Sprechintensität, Phonetica 5, 99 – 115 [40 54 63]
- 701 Zwicker, E. and Feldkeller, R. 1967: Das Ohr als Nachrichtenempfänger, Hirzel, Stuttgart [55]
- 702 Zwirner, E. (Ed.) 1965: Proceedings of the 5th international congress of phonetic sciences at the university of Münster, 16 – 22th August 1964, Karger, Basel [64]

A) Biological aspects**9 phylogeny, primatology**

4 96 257 274 320 322 376 384 387 388 420 427
 484 485 550 567 583 596

10 anatomy

12 33 279 302 376 422 556 619

11 larynx

41 177 191 197 226 279 284 285 286 293 298 321
 337 342 363 371 375 425 442 506 553 585 616 628
 629 630 631 632 696

12 physiology, EEG, EMG, ...

28 72 154 196 200 218 226 227 229 251 270 280
 302 312 324 337 371 374 376 391 420 453 485 492
 553 575 592 613 630 631 643 648 661 674 675 689
 696

13 genetics

42 76 77 120 262 444 452 455 503 623 668

14 age

33 36 37 124 146 153 288 294 295 298 321 328
 432 459 461 491 492 499 519 520 521 564 570 590
 640 679

15 child development

5 6 7 9 23 38 39 100 110 115 139 140
 144 147 155 167 168 172 173 201 202 261 288 290
 291 294 296 310 318 321 329 330 331 335 385 387
 388 396 424 431 435 439 457 467 473 497 498 504
 573 574 627 636 660 666 669

16 sex

16 25 29 39 40 42 53 105 117 118 137 309
 310 313 334 349 368 369 385 392 419 460 550 561
 562 622 640 665 666 690

17 race

27 157 290 291 296 300 311 369 451 568

18 twins

66 67 203 216 217 396 473 551 552

19 constitution

14 181 238 267 325 334 344 345 353 406 422 448
 599

20 voice quality

16 74 95 104 108 128 160 190 200 255 268 280
 287 298 303 307 308 311 336 340 342 363 366 372
 373 391 392 426 436 437 446 454 456 458 464 472
 474 482 499 512 515 518 547 560 566 597 616 639
 661 662 663 672 673 676 679 686 690 692 693 694

B) Medical aspects**21 psychiatry**

11 24 90 91 95 102 255 256 353 418 441 444
 448 451 471 534 591 633 698

22 drug influences

117 118 137 193 239 252 471 595

23 menstruation, birth...

29 40 126 194 195 307 363 423

24 castrato voice

75 271 450

25 other medical aspects

10 11 15 24 25 28 41 55 57 89 119 125
 144 190 199 214 219 222 229 235 242 263 268 277
 308 324 335 340 371 456 458 469 565 574 587 659
 664 665 667 677 678

C) Speech and language sciences**30 languages**

32 43 79 103 114 136 156 157 249 250 270 338
 440 503 505 604 623 644 650

31 speech analysis

5 18 21 27 34 35 47 49 50 51 65 136
 158 161 169 177 189 224 245 248 249 257 270 272
 309 329 330 331 346 359 365 367 402 403 409 440
 443 463 464 476 481 506 531 561 562 604 607 609
 614 618 619 625 626 640 651 658 684 685 697

32 automatic speech recognition

35 78 129 130 133 243 275 314 315 355 357 390
 445 460 475 477 493 494 496 500 501 505 511 516
 517 525 555 557 578 612 656 657 671

33 speech synthesis

21 34 82 445 589 611

34 speaker adaptation

51 209 243 314 315 316 347 358 394

D) Psychology and sociology**40 psychology general**

48 74 87 93 145 158 159 164 165 169 188 206
 236 240 242 343 378 434 450 468 514 549 554 646
 664 668 700

41 personality

2 3 8 16 48 62 81 82 83 97 127 142
 148 163 166 179 182 184 185 247 260 264 273 301
 350 351 408 417 418 438 447 448 449 483 529 530
 537 539 542 544 546 547 566 576 594 621

42 behavioural sciences

4 14 84 96 149 154 197 234 253 274 278 305
 386 389 395 398 402 419 438 439 466 472 487 502
 513 518 536 538 540 545 575 586 591 592 593 594
 639 642 643 644 645 646 647 648 668 670 674 675
 682 683 698 699

43 social aspects

13 79 80 113 152 180 182 223 228 278 305 306
 360 361 362 395 447 454 468 541 543 547 548 563
 650

44 emotions, psychology of expression

108	122	123	128	143	150	167	170	171	183	197	207
208	225	304	320	350	470	514	572	617			

E) Acoustical analysis50 intensity, pitch and F0

23	38	45	52	53	112	115	124	126	139	140	141
149	150	167	168	171	172	173	178	186	202	222	258
261	262	268	276	280	284	285	286	290	291	293	294
295	296	299	300	307	312	316	336	341	348	379	392
406	431	432	435	444	455	465	467	512	533	560	564
582	587	597	610	622	636	654	662	667	680	684	685

51 spectrographic methods

24	36	46	49	60	61	105	116	131	134	153	207
208	219	255	265	277	287	327	330	332	333	354	358
393	399	488	490	531	580	618	637	652	653	676	677
683											

52 average speech spectrum

88	89	98	135	213	231	253	254	256	289	313	338
339	355	404	405	411	471	474	489	595	662	695	

53 general voice acoustics

12	64	83	174	175	176	365	367	532	556	589	628
629	632	641									

54 prosody, speech rhythm

45	52	53	63	72	73	102	112	113	155	159	170
196	234	238	304	316	324	386	403	421	441	462	576
582	697	699	700								

55 auditory analysis

63 69 70 81 85 103 122 123 124 147 160 164
 165 166 178 188 216 218 267 269 297 301 304 309
 364 428 429 430 479 507 522 546 549 554 605 663
 681 701

F) Cetera**60 signal processing**

22 131 134 162 224 231 259 317 319 357 366 379
 380 381 382 407 414 425 443 475 476 495 559 589
 651 652 677

61 statistics

26 30 44 71 92 93 101 111 138 192 380 381
 404 405 458 476 489 490 496 523 524 531 535 548
 558 573 577 578 581 584 588 620 634 638 671 695

62 experimental design

87 147 323 364 437 442 453 463 487 585

63 literature

54 266 372 373 374 375 536 574 629 638 645 669
 700

64 monographs, collections

1 56 161 169 174 175 176 191 228 244 251 370
 397 400 410 445 469 472 532 536 543 626 670 702

65 speaker recognition

17	19	20	30	31	46	58	59	60	61	69	70
71	85	86	88	94	98	99	106	107	109	110	116
121	131	132	134	135	151	153	187	198	204	205	209
210	211	212	213	215	220	221	224	230	231	232	233
241	245	246	254	265	266	281	282	289	292	297	319
326	327	332	333	347	352	356	359	364	380	381	382
383	398	401	402	404	405	412	413	415	416	429	430
433	436	462	463	478	479	486	489	490	502	507	508
509	510	511	526	527	528	557	569	571	577	579	580
606	607	608	615	624	625	635	637	655	657	681	687
688	691	695									

66 handwriting

66	67	68	227	236	237	238	239	240	283	343	377
423	591	598	599	600	601	602	603	698	699		

Anhang

(1) Anthropometrische Merkmale (in eckigen Klammern stehen die Nummern nach Martin/Saller (1959), sonstige Quellenangaben, die Definition des Maßes nach Meßpunkten oder ein arithmetischer Ausdruck, nach dem sich das Merkmal aus anderen Merkmalen berechnet)

Kopfmaße:

- K1 Kopflänge [1.]
- K2 Kopfbreite [3.]
- K3 kleinste Stirnbreite [4.]
- K4 Jochbogenbreite [6.]
- K5 Unterkieferwinkelbreite [8.]
- K6 Gesichtshöhe [18.]
- K7 Nasenhöhe [21.]
- K8 Nasenbreite [13.]
- K9 innerer Augenabstand [9.]

Körpermaße:

- K10 Körpergröße [1.]
- K11 Höhe des Subnasale [3., nach Karolyi 1971]
- K12 Höhe des Suprasternale [4.]
- K13 Höhe des Iliocristale [12.]
- K14 Höhe des Symphysision [6.]
- K15 Schulterbreite [35.]
- K16 Brustbreite [36.]
- K17 Beckenbreite [40.]
- K18 Armlänge [46.]
- K19 Unterarmlänge [48.]
- K20 Handlänge [49.]
- K21 Daumenlänge [Stylien – Daktylion I]
- K22 Handbreite [52.]
- K23 Halsumfang [63.]
- K24 Unterarmumfang [66.]
- K25 Handumfang [67.(2)]
- K26 Fettschichtdicke am Unterarm [nach Knußmann (1968, S. 370)]
- K27 Körpergewicht [71.]

Sekundäre (aus anderen berechnete) Maße und Indices:

- K28 Halslänge [K11 – K12]
- K29 Rumpfhöhe [K12 – K14]
- K30 Beckenhöhe [K13 – K14]
- K31 Androgynie nach Tanner [$3 \cdot K15 - K17$]
- K32 Stammhöhe [K10 – K14]
- K33 Oberarmlänge [K18 – K19]
- K34 Index der Körperfülle [$K27/(K10)^3$]
- K35 Halslänge/ – umfang [K28/K23]
- K36 Unterarmumfang/ – länge [K24/K19]
- K37 Fettschichtdicke/Handumfang [K26/K25]
- K38 Gesichtshöhe/Jochbogenbreite [K6/K4]
- K39 Jochbogen – /Kopfbreite [K4/K2]
- K40 Rumpfhöhe/Körperhöhe [K29/K10]
- K41 Halslänge/Körperhöhe [K28/K10]
- K42 Skeletischer Index [$K14/(K10 - K14)$]
- K43 Armlänge/Körperhöhe [K18/K10]
- K44 Unterarm – /Oberarmlänge [K19/K33]
- K45 Schulterbreite/Körperhöhe [K15/K10]
- K46 Brustbreite/Körperhöhe [K16/K10]
- K47 Beckenbreite/Körperhöhe [K17/K10]
- K48 Handbreite/Handlänge [K22/K20]
- K49 Handumfang/Körperhöhe [K25/K10]
- K50 Kopflänge/Körperhöhe [K1/K10]
- K51 Jochbogenbreite/Kopflänge [K4/K1]
- K52 Beckenbreite – Schulterbreite [K17 – K15]
- K53 Konstitutions – Diskriminanzfunktion nach Knußmann (1968, S. 370)

Fragebogen zur Selbstbeurteilung der Sprechstimme

Sprechprobe Nr. _____

Eigenschaft

Ausprägung

S1	<u>ungezwungen</u>	1	2	3	4	5	6	7
S2	<u>sonor</u>	1	2	3	4	5	6	7
S3	<u>beschwingt</u>	1	2	3	4	5	6	7
S4	<u>dünn</u>	1	2	3	4	5	6	7
S5	<u>überzeugend</u>	1	2	3	4	5	6	7
S6	<u>angenehm</u>	1	2	3	4	5	6	7
S7	<u>kräftig</u>	1	2	3	4	5	6	7
S8	<u>deutlich</u>	1	2	3	4	5	6	7
S9	<u>einfach</u>	1	2	3	4	5	6	7
S10	<u>nüchtern</u>	1	2	3	4	5	6	7
S11	<u>hell</u>	1	2	3	4	5	6	7
S12	<u>aufgeweckt</u>	1	2	3	4	5	6	7
S13	<u>bestimmt</u>	1	2	3	4	5	6	7
S14	<u>ruhig</u>	1	2	3	4	5	6	7
S15	<u>sympathisch</u>	1	2	3	4	5	6	7
S16	<u>klar</u>	1	2	3	4	5	6	7
S17	<u>weich</u>	1	2	3	4	5	6	7
S18	<u>fest</u>	1	2	3	4	5	6	7
S19	<u>beherrscht</u>	1	2	3	4	5	6	7
S20	<u>sachlich</u>	1	2	3	4	5	6	7
S21	<u>natürlich</u>	1	2	3	4	5	6	7
S22	<u>zaghaft</u>	1	2	3	4	5	6	7
S23	<u>klangvoll</u>	1	2	3	4	5	6	7
S24	<u>melodisch</u>	1	2	3	4	5	6	7
S25	<u>frisch</u>	1	2	3	4	5	6	7

Fragebogen zur Beurteilung der Stimme und Sprechweise

Sprechprobe Nr. _____

Beurteiler Nr. _____

Eigenschaft	Ausprägung	subj. Sicherheit
B1 ungezwungen	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B2 sonor	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B3 tief	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B4 dünn	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B5 gekünstelt	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B6 angenehm	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B7 kräftig	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B8 deutlich	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B9 langsam	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B10 nüchtern	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B11 hell	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B12 aufgeweckt	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B13 bestimmt	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B14 ruhig	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B15 sympathisch	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B16 klar	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B17 weich	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B18 fest	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B19 dumpf	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B20 verständlich	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B21 natürlich	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B22 zaghaft	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B23 klangvoll	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B24 melodisch	1 2 3 4 5 6 7	1 2 3 4 5 6 7
B25 frisch	1 2 3 4 5 6 7	1 2 3 4 5 6 7

Fragebogen zur Beurteilung von Sprechern

Sprechprobe Nr. _____

Beurteiler Nr. _____

Eigenschaft

Ausprägung

subj. Sicherheit

a) Körperliche Merkmale

B26	<u>groß</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B27	<u>kräftig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B28	<u>langköpfig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B29	<u>breitschultrig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B30	<u>schlank</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B31	<u>schmächtig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B32	<u>großhändig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B33	<u>langarmig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B34	<u>breitköpfig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B35	<u>schmalgesichtig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B36	<u>langbeinig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B37	<u>dickarmig</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7

b) Psychische Merkmale

B38	<u>extrovertiert</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B39	<u>verschlossen</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B40	<u>unsicher</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B41	<u>zufrieden</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B42	<u>ausgeglichen</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7
B43	<u>gestört</u>	1	2	3	4	5	6	7	1	2	3	4	5	6	7

c) Lebensalter ca. _____ Jahre

Ingo Hertrich

Lebenslauf

- 1955: Geburt am 2. Mai in Stuttgart
Mutter: Heidi Hertrich, geb. Krauß, Ärztin
Vater: Dr. Otto Heinrich Hertrich, Arzt.
- 1959: Umzug nach Bonn
- 1961-64: Besuch der Karlschule in Bonn (Grundschule)
- 1964: Umzug nach München
- 1964/65: Besuch der Grundschule am Mariahilfplatz
(4. Klasse)
- 1965-74: Besuch des Wilhelmsgymnasiums in München (alt-
sprachlich)
- 1974/75: Grundwehrdienst als Sprechfunker (in Murnau und Bad
Reichenhall)
- 1975-80: Studium der Biologie an der Ludwig-Maximilians-
Universität München
- 1977: Diplom-Vorprüfung (allgemeine Biologie, Physik,
Chemie, Anthropologie)
- 1978/79: Studienaufenthalt in England (Cambridge), insgesamt
7 Monate
- 1980: Diplom-Hauptprüfung in Biologie (Anthropologie und
Humangenetik, Zoologie und physikalische Chemie)
- 1981/82: Anfertigen der Diplomarbeit
- 1982: Heirat mit Monika Schrezenmeir (Studentin, Biologie
und Mathematik)
Geburt der Tochter Samira
- 1984: Geburt der Tochter Rosalin
- 1983-86: Promotionsstudium in Biologie an der Universität
München

